

DOCUMENT RESUME

ED 288 896

TM 870 632

TITLE Test Resource Guide, 1987.
INSTITUTION New York City Board of Education, Brooklyn, N.Y. Div. of Special Education.
PUB DATE 87
NOTE 366p.
PUB TYPE Guides - Non-Classroom Use (055) -- Book/Product Reviews (072)

EDRS PRICE MF01/PC15 Plus Postage.
DESCRIPTORS Academic Achievement; *Diagnostic Tests; *Disabilities; Educational Diagnosis; Educational Testing; Elementary Secondary Education; Handicap Identification; *Limited English Speaking; Non English Speaking; *Special Education; Test Interpretation; *Test Reviews; Test Selection; Test Use; *Vocational Evaluation

IDENTIFIERS Education for All Handicapped Children Act; New York City Board of Education; School Based Support Teams

ABSTRACT

The Test Resource Guide (TRG) is designed to provide professionals in the New York City (New York) Public Schools with a compendium of assessment devices for use in decision making; specifically, determining eligibility for special education services, planning an instructional program, and charting progress. The introduction to TRG describes: (1) testing procedures; (2) the psycho-educational model of assessment adopted by the New York City Board of Education; (3) standards for achieving quality assessments; and (4) the uses, development, and organization of TRG. This guide may be used as part of the assessment planning phase to help select appropriate tests, as a tool for increasing professional awareness of available tests across a number of domains, and as a staff development tool. Approximately 340 pages of test reviews and information are presented. Test reviews include the following items: intended purpose of the test, description, norm data, reliability, validity, and a section containing special alerts, effective uses, and comments. A wide range of tests are reviewed: tests of academic functioning (achievement batteries, readiness, reading, language arts, mathematics), adaptive behavior, cognitive functioning, communication and language, perceptual and motor functioning, social-emotional assessment, bilingual-limited English proficiency, and vocational assessment. State and federal laws and regulations are discussed. Chancellor's regulations are appended, as well as a publisher's directory, glossary, and bibliography. (MDE)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED288896

TEST RESOURCE GUIDE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

1987

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Irene Strum

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

TM 870632

NEW YORK CITY BOARD OF EDUCATION

Robert F. Wagner, Jr., President

Irene Impellizzeri, Vice President

Gwendolyn C. Baker

Richard I. Beattie

Stephen R. Franse

James F. Regan

Edward L. Sadowsky

Nathan Quinones, Chancellor

Charles I. Schonhaut, Deputy Chancellor

Edward Sermier, Chief Administrator
Division of Special Education

FOREWORD

This guide has been prepared as part of New York City Public Schools' commitment to quality assessment. Quality assessment is a dynamic process which ensures that decision making occurs in a nonbiased manner leading to the development of a program for meeting individual educational needs. All students, including those who are culturally or linguistically diverse and those who have or may have handicapping conditions, are entitled to assessments with instruments which will report their abilities, skills, and achievement fairly. All professionals engaged in the assessment process and/or who use test results should have easy access to the most current information on tests and measures.

It is our hope that the use of this guide will facilitate the decision-making process in regard to eligibility for special education services, as well as instructional program planning and measurement of pupil progress.

Edward Sermier
Chief Administrator

ACKNOWLEDGEMENTS

This Resource Guide is a project of the Division of Special Education, Edward Sermier, Chief Administrator, prepared under the direction of Gloria Cole, Director, Office of Program Support.

The project was coordinated by Irene Strum, Director, Testing Unit. The format of this Guide was based upon the original work of Aryeh Davidson, Mary-Beth Fafard, Bonnie Fine, Evelyn Goldwasser, Dolores Morris, Olga Romero-Rothenberg, Irene Strum, Elaine Sumaza Smith, and Sharon Stirler Yoshida, under the supervision and guidance of Alan Gartner and Waldemar Rojas.

Elysa Rothenberg served as general editor of this new edition under the direction of Irene Strum. Sara Fischer and Robert Margolis also assisted in test reviews and editing.

The original Guide's test reviewers were Gertrude Ashur, Dominick . Balducci, Claire Barabash, Jerry Baraz, Lawrence Becker, Rita Carroll, Raymond DeBold, Jeffrey Glass, Isabel Gonzalez, Ann Gorewitz, Henry Lipton, Sidney Marks, Charles Miller, Jack Ness, Howard Nissman, Victor Ortiz, Sal Pappalardo, Esther Perez, Sidney Rabin, Sol Ribner, Evelyn Rojas, Florence Sapiro, Donald Weidis, and Michael Zangwill. Additional tests for this edition were reviewed by Richard Balukas, Dom Capone, Judy Dumas, Sarina Gwartzman, Judy Litvak, Bennet Rackman, Sol Ribner, David Schultz, Hilda Sorell, Nancy Taylor, and Pat Valentino.

With the cooperation of Norman Fredman at Queens College and Laura Harckham at Fairleigh Dickinson University, the following graduate students also reviewed tests for this edition: Roberta Ackerman, Jeffrey Adler, Judith Alberts, Jenni Conn, Dominic Cuoccio, Kathleen Faabri, Carol Gaskill, George Goldschlag, Rochelle Goldschlag, Janet C. Gresco, Laila Herstol, Susan Hindes, Jules Levy, Beth Maravnick, C.J. Mazzarella, Rosemary McKay, A. Meo, Halycan Nicome, Helene Rauch, Richard Rubin, Rhoda Sherry, Sheila Simanowitz, Ellen Tauber, Mary Ellen Vastov, M.E. Vogel, and Jane Wolfe

The text for this edition was typed by Karen Sanchez, Julia Sedito, and Thelma Sternstein.

We are grateful to all the staff who contributed their time and energy to improve non-biased assessment practices for New York City's children.

TABLE OF CONTENTS

Introduction	1
Test Reviews	
Academic Functioning	13
Achievement Batteries	15
Readiness	35
Reading	45
Language Arts	71
Mathematics	79
Adaptive Behavior	99
Cognitive Functioning	141
Communication and Language	187
Perceptual and Motor Functioning	225
Social/Emotional Assessment	245
Tests for Bilingual and/or Limited English Proficiency (LEP) Students	275
Vocational Assessment	311
Glossary	359
Appendices	
A. The Law and the Assessment Process	365
B. Chancellor's Regulations on Non-Discriminatory Assessment	369
C. Materials for Updating the Guide	381
D. Approval Form	383
E. Publishers/Distributors Directory	385
Bibliography	391
Test Index	399

INTRODUCTION

The Test Resource Guide is designed to provide professionals in the schools with a compendium of assessment devices for use in decision making. The need for this Guide stems directly from the goals of assessment (Oakland, 1977): determining eligibility for special education services, planning of an instructional program, and the charting of progress. Norm referenced and criterion referenced tests are useful in the school setting to the extent that they address these goals and facilitate appropriate instruction for all students. Projective techniques add descriptive information about a student's feelings, emotions, conflicts, and problems which may have an impact on results obtained from other instruments and procedures.

A premise underlying this Guide is that tests and test data are only as good as the skills and expertise of the people who use them. The information in the Guide will add to the knowledge and skills of New York City staff. In addition to knowledge about standardized tests and formal testing, test users need skills in observation, interviewing, working with parents, and decision making, as well as knowledge of student behavior, cultural, and linguistic factors.

The decision to use a specific instrument depends upon the needs of the individual student. Since an individual test or technique yields limited information, data from multiple sources must be considered when making placement and programmatic decisions. Consequently, no one test or technique should ever be used or interpreted in isolation from other assessment information.

ASSESSMENT AND SBST/CSE

Assessment is a process of measuring a student's present level of performance, identifying a student's strengths and weaknesses, and monitoring a student's progress.

In their evaluation of a student, the School Based Support Team/Committee on Special Education (SBST/CSE) must ensure that the following mandated practices have been implemented (PL94-142, 1975; Regulations of the Chancellor, A831, 1984).

1. Tests and other assessment procedures:

- a. Are provided and administered in the student's dominant/native language, if this is not English, and English or other mode of communication.
- b. Have been validated for the specific purpose for which they are used.
- c. Are administered by trained personnel in accordance with the instructions provided by those who developed such tests or procedures.

2. Tests and other assessment procedures include those tailored to assess specific areas of educational need and not merely those which are designed to provide a general intelligence quotient.
3. Tests are selected to ensure that when a test is administered to a student with impaired sensory, manual, or speaking skills, the test results accurately reflect the student's aptitude or achievement level or whatever other factors the test purports to measure, rather than reflecting the student's impaired sensory, manual, or speaking skills, except where those skills are factors which the test purports to measure.
4. No single procedure is used as the sole criterion for determining an appropriate educational program for a student.
5. The evaluation is made by a multidisciplinary team or group of persons, including at least one special educator.
6. The student is assessed in all areas related to the referral.
7. The evaluation includes at least one observation of the student in the primary instructional setting.
8. Parent/guardian interviews are conducted in the parent's dominant language. Translators shall be used when needed.

Psychoeducational model of assessment. The psycho-educational model of assessment, which has been adopted by the New York City Board of Education, emphasizes the individual approach to assessment. There is no established battery of tests appropriate for all students. For each student, the clinician has the responsibility to determine initial and ongoing eligibility for special education service and to identify specific strengths and weaknesses of the student in the school environment so that appropriate instructional methodologies and other interventions can be designed and implemented.

Psychoeducational assessment is situation specific. The decision for selecting an area of assessment depends upon the reasons for referral and, in reevaluations, upon the student's educational needs. Most students require examination of their cognitive development and achievement in specific curriculum areas. Some students need to be assessed for motor and psychomotor development. Some students need to be assessed in the social/emotional domain. Some students need an in-depth assessment of speech, language, and communication skills.

In addition to areas for assessment, the selection of specific instruments is a decision that should be made for each individual student. Specific instruments are selected in relation to the

individual student being assessed. Sometimes informal procedures including observations precede actual test selection so that the appropriate instruments may be chosen for assessment. Many tests are administered to students of specific ages and are not available for all students. Some tests are recommended for students of low or high ability, for students with high verbal skills, for students with little oral language skills, etc.

Quality assessment. Quality assessment ensures decision making which will meet the needs of each individual. The assessment develops from the referral and proceeds through a series of steps that allows for flexible examination tailored to each student. The reason for referral, the core of the assessment, provides the foundation of the process.

Standards for achieving quality assessment have been identified from the literature to ensure the best professional practices (Morris & Strum, 1986; Strum, 1985). Ten standards of quality assessment are:

1. Integrating the reason for referral into the assessment process. Evaluation of referral data provides the framework for the assessment process. Clarifying the referral should include conferences to pinpoint the academic and/or social behavior(s) of concern (Ysseldyke & Thurlow, 1984). An integral part of this consultation is determination of teacher-pupil interaction in the classroom, scope of curriculum, attitude and expectations of school personnel and other students (Winters & Easton, 1983).
2. Considering linguistic and cultural needs. The validity of the process of assessment demands that the linguistic and cultural background of the students be integrated into the decision making. The ultimate goal of quality assessment is to implement techniques and instruments which aid in the evaluation of the students within the cultural setting in which they grew up and developed. Cognitive and affective development and learning strategies are influenced by the socialization practices of the primary culture (Feuerstein, 1979; Ramirez & Castaneda, 1974).
3. Involving the parent and student in the assessment process. A major goal in quality assessment is to develop a working relationship between parents and student and the SBST. Parents should be fully informed of their legal rights to service and education (PL 94-142, 1975). The achievement of collaboration requires the skillful application of clinical interventions and knowledge (Coughlin & Mears, 1984; Ysseldyke & Thurlow, 1984).
4. Planning and selecting areas of assessment and techniques that match student needs. A multidisciplinary team formulates an assessment to fit the student's need. Assessment measures should be selected on the basis of their educational relevance and

utility (Messick, 1980; Reschly, 1984; Snow, 1984). Careful attention must be paid to the psychometric characteristics of instruments and techniques to ensure that valid and reliable tools are used for the purpose for which they are designed (APA, AERA, & NCME, 1985; Oakland & Matuszek, 1977; Shepard, 1983; Ysseldyke, Algozzine, Regan, & Porter, 1980).

5. Administering instruments/implementing techniques according to guidelines. Concepts underlying assessment procedures and techniques must be ethical, systematic, and accurate. The principles of professional practice and the appropriate considerations regarding the use of standard and non-standard techniques should always apply (Bennett, 1981; Tucker, 1980; Ysseldyke, Algozzine, Regan, and Potter, 1980; Ysseldyke and Regan, 1980; Ysseldyke and Thurlow, 1984). Procedures for data collection must include advance preparation of examiners, mastery of the manual and the test, and selection of a site with appropriate testing conditions (Anastasi, 1982; Cronbach, 1983).
6. Scoring and interpreting data in keeping with professional standards. Reporting scores and interpreting normative scores should be made from instruments that are reliable and valid and have norms appropriate for the population. Appropriate analysis of instruments and techniques is necessary to result in valid and reliable decisions. Technical skills in scoring and interpretation are necessary to provide and maintain quality services (Bennett, 1981). Scoring must be implemented following directions in test manuals (APA et al, 1985; McDermott, 1981).
7. Writing comprehensive reports relevant to educational planning. Written reports provide a means for organizing thinking. The fundamental purpose of reports is documentation of the data gathering activities with the student and parent. Reports serve as guides to the educational decision making process and programming. The report should be cohesive and informative, (Byrne et al, 1977; Wilson, 1980). Scores should be reported when tests are valid, reliable, and have representative norms. When appropriate test norms are not available, the information collected should be of a descriptive nature (Flegenheimer, 1985; Regulation of the Chancellor A831, 1984).
8. Determining the best educational program in the least restrictive environment with access to the full range of educational services. The data collected are used as the basis for decision making (Ysseldyke & Regan, 1980). Systematic comparison of these data to the characteristics of handicapping conditions determines eligibility for special education services. Program placement and services recommended are based on the educational needs of the individual students regardless of their handicapping condition.

9. Developing instructionally relevant individualized education programs. The Individualized Education Program (IEP) is the tool by which parents and team members collaborate to plan an effective program. As a legal document it structures educational goals, promotes communication, and permits accountability (Felker, 1977; Murray, 1980; PL 94-142, 1975; Winters & Easton, 1983). It is completed by converting assessment data into measurable educational goals and objectives. It must be a definitive statement regarding classification, educational programs and related services.
10. Managing the assessment in an effective manner. The management of the quality assessment is determined by application of the aforementioned standards in an ethical, legal, knowledgeable, and skilled manner by all those involved in the process. Factors that are critical to the effective management of assessment practices include technical adequacy of the measurement devices, ease of administration, time efficiency, extent to which practices result in accurate decisions, the degree of intrusion in the student's life, cost effectiveness and the extent to which participants value the assessment process (APA et al, 1985; Ysseldyke & Thurlow, 1984).

USE OF THE GUIDE

The Guide may be used in at least three ways. These are:

As part of the assessment planning phase to help select appropriate tests. To this end, the test descriptions include data on purpose, standardization, reliability, validity, and special alerts. This information, used in conjunction with test manuals, will help in the initial selection of the appropriate tests for an individual student. Particular attention should be paid to the special alerts section. Tests, of course, should be used only for purposes for which they will provide valid and reliable information about the student being tested.

As a tool for increasing professional awareness of available tests across a number of domains. The Guide will provide staff with information about instruments appropriate for assessing students with special needs and diverse cultural and linguistic backgrounds.

As a staff development tool. The Guide will provide staff with material for discussion and professional growth as part of on-going training.

DEVELOPMENT OF THE GUIDE

In the Spring of 1982, the Office of Program Development conducted a survey of CSE Chairpersons, Clinical Supervisors, and SBST members to determine test availability, kind of test information needed, and type of format the Guide might take to be most appropriate for its users. A committee of experts in the area of assessment then identified specific criteria for reviewing tests which were incorporated into a checklist.

This checklist consisted of criteria which had been developed by professional organizations and leading experts in the field of assessment (Oakland, 1977) and which reflected the legal and technical criteria mandated by federal, state, and local laws and litigation. Tests were recommended for specific purposes on the basis of the requirements of P.L. 94-142 (12A 532, 1975) that "tests be validated for the specific purpose for which they are used", the standards prepared by a joint committee of the American Psychological Association (APA), the American Educational Research Association (AERA), and the National Council on Measurement in Education (NCME) in the Standards for Educational and Psychological Tests (1974), and Part 200 of the New York State Regulations.

For the first edition of the Guide, CSE Chairpersons, Clinical Supervisors, SBST members, and Central staff reviewed tests using the checklist. Tests selected for review were those mentioned as being used by psychologists and educational evaluators; additional tests which were available were also included.

The new 1987 revised edition of the Test Resource Guide was developed because of the publication of many new tests for assessing student eligibility for special education and for instructional planning. Moreover, the importance of criterion referenced tests in instructional planning necessitated the inclusion of these instruments alongside norm referenced tests rather than as a separate volume.

ORGANIZATION OF THE GUIDE

The present Test Resource Guide consists of one volume which includes various types of instruments used in assessment that have been recommended for use with most students in New York City Public Schools; i.e., norm referenced and criterion referenced tests and projective techniques.

To assist the test user in selecting the tests or techniques needed for an individual student, tests are organized into the following categories: academic functioning, which is further subdivided into reading, math, language arts, readiness skills, and general achievement batteries, adaptive behavior, cognitive functioning, communication and

language, perceptual and motor functioning, social/emotional assessment, tests for bilingual students and/or students with limited English proficiency (LEP), and vocational assessment. Within categories or subtopics, tests are arranged alphabetically by title.

The extent of the review on technical development, i.e., norms, reliability, and validity, is based upon type of instrument. These data are crucial for using results from norm referenced tests, whereas they are somewhat less important for criterion referenced tests where performance level is not generally reported. For the latter, information on test administration is most necessary for the examiner.

Norm referenced tests. Norm referenced tests (NRT) compare an individual's performance with the performance of his/her peers. Raw scores are converted into standard scores, such as an IQ, or derived scores, such as percentiles and grade equivalents, which were determined by the results obtained from the standardization sample. Scores may be interpreted meaningfully for students for whom the standardization sample is representative. The Chancellor's Regulation (A831, 1984) permit reporting of scores only where tests were normed on populations representative of the student.

Criterion referenced tests. Criterion referenced tests (CRT), on the other hand, report an individual's level of mastery in a specific subject, domain, or area. They provide information on the extent of knowledge or skill a student has in a specific area and are, therefore, useful for making instructional planning decisions.

The differences between NRTs and CRTs are summarized in Figure 1 on page 8.

Projective techniques. Projectives are used by clinicians to obtain descriptive information about an individual's personality. Projectives may provide insight into the psychodynamic interplay among intellectual and emotional factors as they influence an individual's adaptation, problem solving, and thinking processes. This insight results from qualitative or descriptive interpretation of responses rather than from quantitative data. It has been strongly advised that projectives be used cautiously in schools (Noll et al, 1979). Most frequently, this category of test is selected to assist in answering questions about fantasy, motivation, and reality testing. These tests should only be used when there is a particular basis to believe that such an instrument is appropriate to the individual student.

Figure 1
Comparison of NRTs and CRTs

Factor	<u>NRT</u>	<u>CRT</u>
Test Items	test items are constructed to <u>discriminate--easy for some, hard for others</u>	items are constructed to test whether student <u>can perform</u> the skill specified by the objective
Timing	usually a <u>timed</u> test	an <u>untimed</u> test
Directions	<u>formal</u> test - test directions and examples cannot vary	<u>Informal</u> test - teacher can vary directions and provide additional examples
Test Level	student takes test book on <u>specified grade level</u>	student takes booklets according to <u>functional level</u>
Scores	<u>compares</u> student performance to a set of national <u>norms</u>	indicates student performance relative to <u>specific objectives and content</u> . Results may be translated into IEP goals.
Interpretation	indicates <u>age/grade equivalent</u> and percentile	indicates whether student <u>does</u> or <u>does not know</u> objectives.
Test Taking Instruction	teaching for test invalidates test	teaching for the objectives is desirable and expected

Source: Teichert & Daniel, 1980

SPECIAL ALERTS/EFFECTIVE USE/COMMENTS SECTION OF THE GUIDE

The Special Alerts/Effective Use/Comments section is the most pertinent and practical aspect of each review, especially when the user is selecting tests. Each section includes the following areas: (1) use of test results in educational decision making; (2) appropriate populations; (3) limitations on the interpretation of test results; (4) additional considerations relating to practice.

1. Use of test results. Each of the tests used in the Guide provides information which may be used to assist in various educational decision making processes. These include:

Screening. A test may be used to identify students who may differ sufficiently from their peers in a specific area to warrant further diagnostic assessment. These tests are generally group administered examinations. They must have high content validity; reliability estimates of at least .80 must be reported in the test manual, technical bulletin, or additional research (Salvia & Ysseldyke, 1985).

Determination of performance. A test may be used to measure ability, functional level, and age or grade level in a particular area. Most of these tests are individually administered; however, some standardized achievement tests which report grade levels may be group administered. Most of these tests are norm referenced. Descriptions of standardization samples and content and criterion related validity are especially important for tests yielding this information. Moreover, reliability estimates of .90 or higher are necessary if test scores are to be used to determine individual performance (Salvia & Ysseldyke, 1985). Nunnally (1978) has recommended .95. Results from tests which do not present such levels of reliability should be clearly qualified as to their influence and impact on placement/program decision making. Few CRTs may be used for this purpose.

Instructional planning. A test may be used to assist the user in determining what and how to teach a student. Few NRTs are applicable for this purpose because of the limited number of items included per domain and their questionable reliability (Salvia & Ysseldyke, 1985). CRTs are generally used for planning instruction; they generally provide short- and longterm objectives which may be included on the IEP.

Evaluation of student progress. A test may be used to monitor student growth over a period of time. Most NRTs and many CRTs are useful for this purpose if their content is representative of the curriculum; (i.e., the test has content validity) and their reliability is greater than .90 for individual comparison or .60 for group comparisons (Salvia & Ysseldyke, 1985). Although this Guide does not address the use of tests for program evaluation (i.e., special projects, grants, curriculum innovations, etc.), tests selected for this purpose should have technical adequacy as described above.

Most educational decision making processes are useful for individual classification and placement decisions. Classification and placement decisions are based on a team decision using multiple sources of information. These decisions are one of the results of the assessment process. In addition, NRTs, CRTs, and

projectives may be used to provide supplementary, descriptive information about the way the student responds to the task demands of the test without reference to a norm. No standard criteria (e.g., reliability and validity) are required for tests and techniques used in this manner (Anastasi, 1976).

2. Appropriate population. Based on the technical adequacy of the test and the diversity of the New York City school population, recommendations are made concerning the appropriate use of the test with respect to specific populations (e.g., age, grade, handicapping conditions). Reference is also made to the possible use of the test with students of Limited English Proficiency (LEP) and students of culturally and linguistically diverse backgrounds. Recommendations for reporting scores for bilingual and/or LEP students are also indicated.
3. Limitations on the interpretation of test results. Specific cautions regarding the use and interpretation of test results reflect technical aspects of the test and current research regarding particular test use. With respect to limitations on use, the practitioner should examine original sources as well as the test manual for the details on limitations.
4. Additional considerations relating to practice. Comments regarding practical use of the test are provided to assist the user in selecting a test.

TESTS NOT REVIEWED IN THE GUIDE

Users are invited to report information as to the usefulness of material in this Guide and suggestions as to additional information for future editions.

Tests currently in use but not reviewed in the edition may be used to provide supplementary, descriptive information (see above). These tests and newly developed tests should be submitted by the CSE Chairperson to the Office of Educational Assessment for review. The test should be submitted as indicated in Appendix C. The submitted test will be reviewed and results reported according to the publication updating schedule.

UPDATING THE GUIDE

A joint committee of the APA, AERA, and NCME revised the standards for testing, giving specific attention to linguistic and cultural differences of students. This new edition of the Test Resource Guide has incorporated the new standards and criteria (APA et al, 1985).

As additional tests and procedures are recommended for review by Guide users, research studies provide evidence for support of specific instruments and procedures, and test users provide feedback on specific instruments, supplementary material will be provided for inclusion. A notification of the new schedule for adding materials to the Guide will be issued regularly.

MONITORING TEST USE

It is the responsibility of the Office of the Chief Administrator of the Division of Special Education to monitor closely test usage based on the information contained in this Guide.

Whenever tests are used beyond their stated intended purpose, as indicated in the Special Alerts/Effective Use/Comments section: (1) written approval must be obtained from the CSE Chairperson or Clinical Supervisor, (2) such use must be noted in the student's record, and (3) extraordinary care must be taken in the use of data derived from the instrument. An outline of the material necessary for recording such approval is contained in Appendix D.

This procedure is intended to insure that the evaluation safeguards detailed in P.L. 94-142 are implemented.

ACADEMIC FUNCTIONING

Tests of academic functioning are organized into the following subcategories: achievement batteries, readiness, reading, language arts, and mathematics. As academic achievement tests, these tests measure a student's proficiency in academic areas. They should reflect the curriculum; i.e., have content validity. Some achievement tests are multiple skill batteries while others are specific to one content area, such as reading or mathematics.

Achievement tests are norm referenced, criterion referenced or both. Those that are norm referenced permit the examiner to measure the extent of student learning in comparison to other students and thereby report a performance level. CRTs may provide extensive information for instructional planning because of their indepth analysis of an individual's strengths and weaknesses in a specific domain.

Test findings should be interpreted taking into account the influence of cultural and linguistic differences.

ACHIEVEMENT BATTERIES

Tests included in this subcategory are multiple skill examinations which assess student achievement in more than one area. Most of them estimate a student's proficiency in reading and mathematics.

Title	<u>Basic Achievement Skills Individual Screener (BASIS)</u>	Grade	<u>1st - 12th</u>
Author	<u>J. Sonnenschein</u>	Time	<u>Untimed (one hour)</u>
Copyright Date	<u>1983</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

BASIS is designed to measure reading, mathematics, spelling, and writing achievement, providing both norm referenced and criterion referenced information in these areas.

Description

BASIS is an individually administered test consisting of oral reading, printed computation, dictated word problems and spelling, and an optional writing sample. Reading passages contain a cloze exercise and are additionally scored using miscue analysis.

Norm Data

The standardization sample consisted of 3,296 students in grades 1 through 12 in 66 school systems from 23 states. Geographic distribution, socioeconomic status, and ethnic composition were similar to the national population. Approximately four percent of the students were identified as disabled; however, these were mainstreamed students, not students in self-contained special education classes.

Reliability

KR 20 estimates of internal consistency for each school grade and full-year age group were generally above .90. Test-retest reliability coefficients for selected grades were considerably lower; only two coefficients obtained for a group of 60 second graders were above .90. Several reliability studies conducted among small samples of special education students in self-contained classes yielded reliability coefficients generally above .90 in reading, math, and spelling.

Validity

Content validity was established through test development procedures. The *p* values indicated the accuracy of the grade-level placement of items. Correlations with report card grades, national achievement tests, and statewide assessments ranged between .25 and .72. Validity studies among small groups of special education students in self-contained classes resulted in correlations generally of .40 or higher of BASIS with CAT.

Special Alerts/Comments

The information obtained from BASIS may be useful in determining performance levels in reading, math, and spelling. This information may also be useful for making screening, instructional planning, including development of short- and longterm IEP objectives, and evaluation of student progress decisions, especially above readiness levels. There is a high degree of correspondence between test items and the MTEs.

Criterion referenced test information is obtained during test scoring. Information for diagnosing reading needs is based on the psycholinguistic model of miscue analysis. Although directions for miscue analysis are included in the manual, only evaluators who have been trained in miscue analysis will be able to provide detailed instructional recommendations on the IEP.

While norms are not representative of students with handicapping conditions, the scores indicate the achievement of a student in relation to other students of similar age/grade. Caution should be exercised when early childhood students and recently arrived LEP students are included in the testing population, since the degree of exposure to English language skills influences performance. Scores may not be reported for bilingual and/or LEP students.

Reference

Manual

Publisher

The Psychological Corp.

Title	<u>California Achievement Tests (CAT)</u>	Grade	<u>K - 12th</u>
Author	<u>None</u>	Time	<u>Varies according to level</u>
Copyright Date	<u>1978</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The CAT is an achievement test providing information on pre-reading, reading, spelling, language, mathematics, and reference skills.

Description

The CAT is a group administered test consisting of multiple choice questions. There are ten levels. Except for the first three levels, each level is available in Forms C and D. The reading test provides scores on reading vocabulary, reading comprehension, and total reading. The pre-reading test, level 10, consists of scores for Alphabet Skills, Letter Sounds, Visual and Auditory Discrimination, Total Pre-reading and Mathematics. Other scores resulting from administration of the complete battery are: Spelling, Language Mechanics, Language Expression, Total Language, Math Computation, Math Concepts, Total Math, Reference Skills.

Norm Data

The CAT was standardized on a sample of 200,000 students from kindergarten through twelfth grades attending public and Catholic schools. For the public schools, a random stratified sample was obtained on the basis of geographic region, size and enrollment, and general achievement.

Reliability

Kuder-Richardson formula 20's (KR20) and standard errors of measurement were calculated for standardization sample subtest and total scores. KR20's were generally reported above .80; however, for some subtests, somewhat lower indices were indicated. Total score test-retest reliability based on 2-to 3-week intervals was .75. The reliabilities for individual subtests ranged from .23 to .81, with most in the .60 to .70 range.

Validity

Test development practices were described in detail to support content validity. Test items were reviewed to minimize racial, ethnic, and sex bias. The reading subtests are valid for New York City since the items reflect New York City's objectives and curricula.

Special Alerts/Comments

The results from the CAT may be useful for screening and evaluation of student progress decisions, with respect to pre-reading, reading, spelling, language, mathematics, and reference skills. There is a high degree of correspondence between test items and the MTEs.

Given its group administration and scoring procedures, the information obtained from the CAT is insufficient for decisions regarding individual instructional planning and determination of performance levels for students who may have handicapping conditions. For these students, further individual assessment is necessary.

Although the norms are not representative of students with handicapping conditions, the scores indicate the achievement of a student in relation to other students of similar age/grade.

Caution should be exercised when early childhood students and recently arrived LEP students are included in the testing population, since the degree of exposure to English language skills influences performance.

References

Manual
Salvia & Ysseldyke, 1985

Publisher

CTB/McGraw Hill

Title	<u>Comprehensive Inventory of Basic Skills (CIBS)</u>	Grade	<u>K-9th</u>
Author	<u>A. Brigance</u>	Time	<u>Untimed</u>
Copyright Date	<u>1983</u>	Type	<u>CRT</u>

Intended Purpose of Test

The CIBS is designed to assess basic academic skills of students in the elementary and middle schools, especially those students with special needs requiring individualized instruction. It includes the processes of assessment, diagnosis, record keeping, and instructional planning.

Description

The CIBS consists of placement tests and skill assessments similar to its predecessor, the Diagnostic Inventory of Basic Skills (DIBS). The CIBS, however, is more comprehensive; it has been extended to include skills normally taught through grade 9 and also additional skill sequences in readiness, reading, language arts, and math. Four new categories were added: speech, listening, percent, and metrics. Skills originally categorized under vocabulary in the DIBS were reclassified and incorporated into the listening and reading sections. The assessment procedures for the word recognition and reading comprehension tests follow conventional informal reading inventories more closely than the DIBS by making provisions for diagnosing oral reading and silent reading comprehension errors.

The over-all design, the record keeping components, the flip-over manual, and the contents of the examiner's page are similar. The newer manual, however, provides more information and suggestions regarding observations, adaptations, supplemental assessments, related skills, diagnosing difficulties, and references. The CIBS also offers alternate test forms A and B for fifty-one skills.

Skills tested include the following:

Readiness

color recognition, visual motor, visual discrimination, visual memory, gross motor, identification of body parts, directional/postional concepts, fine motor, articulation of sounds, personal data response, alphabet, numeral recognition, number recognition and comprehension, sets, recognition of lower case letters and upper case letters, letter printing, response to books, quantitative concepts, self help

Speech	syntax and fluency, response to a picture, articulation of initial sounds and final sounds
Reading	word recognition, oral reading comprehension, vocabulary, word analysis, auditory discrimination, consonants, vowels, digraphs, blends, diphthongs, suffixes, prefixes, syllabication
Language Arts	spelling, reference skills, listening, writing, graphs and maps
Math	number, number facts, computation of whole numbers, fractions and mixed numbers, decimals, percents, metrics, math vocabulary

Mastery level criteria vary from 67% (e.g. listening vocabulary) to 100% (e.g. alphabetizing skills), depending upon the nature of the skill; some skills are left to the examiner's discretion.

Test Administration

The CIBS is too extensive to be administered in its entirety. The examiner selects the appropriate test depending upon the purpose for assessment (screening, skill level for placement, IEP development) and upon relevancy to student and program needs. Most subtests are administered individually.

The author suggested beginning assessment one grade level below the student's anticipated grade level performance to assure success. The CIBS does not require specialized training in testing. Many of the subtests can be group administered with written responses. No special supplies are needed. A color coding system is recommended as a way of developing on-going assessment data for each student in his/her record book.

Test Information

The CIBS was developed as a more comprehensive assessment instrument than the DIBS. The field test edition was based upon the experience and feedback received from professionals using the CIBS over a period of six years and the author's knowledge, skills, and resources gained from developing his other inventories. The CIBS bridges the skills assessed by the Inventory of Early Development (IED) and the Inventory of Essential Skills (IES).

Content validity was developed by submitting the field test edition for critical evaluation to 141 regular and special educators in 27 representative states of the U.S. and British Columbia. The author also stated that more extensive referencing with text books was used with the CIBS as compared to the DIBS in order to validate the scope and sequence of skills taught.

A reliability study, conducted in New Orleans, correlating forms A and B of eight placement tests, resulted in correlations ranging from .71 to .96. Low reliabilities were attributed to restricted variance, a characteristic of criterion referenced tests (Enright, in press.) Another study is being conducted among special education students in New York City.

No information was presented on how mastery was determined.

Effective Use/Comments

The information provided by the CIBS is sufficient for developing short- and longterm objectives in readiness, reading, oral and written language skills, and mathematics. The information is not sufficient for making screening and performance level decisions.

The test follows the sequence of the METs. While instructional objectives are specified, the manual does not provide strategies for classroom instruction. However, the CIBS provides a comprehensive and effective instructional management system in the basic skills.

Some of the readiness and math tests may be useful in the assessment of linguistically and culturally diverse students. Other subtests must be carefully selected and interpreted with caution for linguistically and culturally diverse students and LEP students.

Reference

Manual

Publisher

Curriculum Associates

Title Inventory of Essential Skill (IES) Grade 7-12
Author A.H. Brigance Time Untimed
Copyright Date 1981

Intended Purpose of Test

The IES is designed to assess basic skills identified as essential to functioning successfully as a citizen, consumer, worker, and family member.

Description

The IES, published in a sturdy looseleaf binder, consists of 2 major components: (1) basic academic skills and (2) applied skills. Each component contains orally administered tests and rating scales on the following:

Basic Academic Skill

word recognition for grade placements, oral reading, reading comprehension, word recognition of functional words, word analysis, reference skills, schedules and graphs, writing, personal data forms, spelling, math grade placement, numbers, number facts, computation of whole numbers, fractions, decimals, percents, measurement, metrics, and math vocabulary.

Applied Skills

health and safety, vocational, money and finance, travel and transportation, food and clothing, oral communication, and telephone skills

Two tests are available for assessing most skills. The criterion for mastery varies according to the test and is explicitly stated on the assessor's page. The grade levels in which skills are normally introduced, taught, or achieved by the student are indicated with each listed skill in the assessor's portion of the inventory.

The test binder may be used as an easel so that the student's page is oriented toward the student when facing the assessor during oral one-to-one testing. Student pages may be duplicated for group administration. No additional materials are needed.

The rationale, specific information regarding assessment procedures, and interpretation of results for each subtest are included. A Student Record Book is provided for recording results.

Test Administration

The IES is too comprehensive to be administered in its entirety. The test administrator selects the subtests to be administered based upon the purpose for testing. Training is required for administering the oral

reading tests. Some tests, however, may be given by a trained paraprofessional. The nature of the test determines whether oral or written responses are required. Most subtests are administered individually.

Technical Information

The IES was developed to provide an assessment program, similar to the K-6 DIBS which is appropriate for order students. It was piloted among student with special needs. More than 100 teachers evaluated the material and assessed the jcontent validity of each subtest. Items which were identified as prejudicial or stereotypical of any segment of the population were deleted.

The IES has content validity through its test development procedures. Reliability was not addressed. Criteria for determining mastery were not presented.

Effective Use/Comments

The information provided by the IES is sufficient for developing short-and longterm objectives on the IEP in the areas tested. The information is not sufficient for making screening and performance level decisions. This test closely follows the sequence of thw MTEs. However, solving mathematical verbal problems was omitted.

While the instructional objectives are specific, the manual does not provide strategies for classroom instruction.

The reading passages, which were adapted fom newspaper articles, should be appealing to most adolescents. The format, however, may not be useful for obtaining information about the mastery of skills for students with learning disabilities.

Most of the subtests may be administered with linguisticlaly and culturally diverse students the LEP students, but the results should be interpreted cautiously.

References

Manual

Publisher

Curriculum Associates

Title	<u>Kaufman Test of Educational Achievement (K-TEA)</u>	Grade	<u>1st - 12th</u>
Author	<u>A. S. Kaufman & N. L. Kaufman</u>	Time	<u>Untimed (up to 1 hour)</u>
Copyright Date	<u>1985</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The K-TEA is intended to measure school achievement in the areas of Reading, Mathematics, and Spelling.

Description

The K-TEA is available in two forms: a Brief Form and a Comprehensive Form. The Brief Form is an NRT, providing scores in Reading, Mathematics, and Spelling; the Comprehensive Form provides NRT and CRT information in five areas: Reading Decoding, Reading Comprehension, Mathematics Applications, Mathematics Computation, and Spelling.

Each kit includes a manual, an Easel-Kit which is a flip book with sequenced test items, Individual Test Records, and Reports to Parents. The Comprehensive Form provides an Error Analysis Summary to categorize student errors.

The K-TEA is individually administered by a professional who is familiar with test procedures and scoring. Subtests in both forms should be followed in their established order.

Norm Data

Standardization was based on two groups tested with the Comprehensive Form in the Fall or Spring of 1983 and 1984. More than 1,000 students in grades 1 through 12 who resided in 16 states were tested at each time period. The samples were stratified by grade, sex, geographic region, socioeconomic status, and race or ethnic group and were representative of U.S. population as reported in the Census for 1983 and 1984.

Norms were developed for the Brief Form by equating with the Comprehensive Form.

Results are reported in terms of raw scores, standard scores, stanines, NCEs, percentiles, age/grade equivalents, and a Battery Composite. All standard scores on both forms have a mean of 100 and a standard deviation of 15. Subtests may be compared in terms of standard score differences.

Reliability

Split-half reliability of the three subtests in the Brief Form, based on the Rasch-Wright model, ranged from .85 to .91; for each age and grade level, coefficients ranged from .72 to .97. Reliability of the five subtests in the Comprehensive Form ranged from .90 to .95; for each age and grade level, coefficients were above .83. Test-retest reliability of the Brief Form, based on retesting 153 students, grades 1 to 12, ranged from .84 to .90. Reliability of the Comprehensive Form, based on retesting 172 students, grades 1 to 12, ranged from .83 to .96. Alternate form reliability was above .80.

Validity

Content validity was developed through test construction and item selection using the Rasch-Wright model. Construct validity was indicated through internal consistency and the relationship of scores to age. Correlations between subtest and Battery Composite scores ranged from .60 to .93 on the Brief Form and from .70 to .93 on the Comprehensive Form.

Concurrent validity was indicated by comparisons with several tests such as the WRAT, PIAT, K-ABC, and PPVT-R; correlations with each ranged from .22 to .92.

Special Alerts/Comments

The K-TEA provides information on school achievement in the areas of Reading, Mathematics, and Spelling which may be used for making performance level, instructional planning, and evaluation of student progress decisions. The Brief Form is especially useful for screening purposes; the Comprehensive Form may be considered a CRT as well as an NRT in that it offers in-depth analysis of academic strengths and weaknesses in order to identify areas needing remediation and provides information for developing short- and longterm objectives on the IEP.

The manual cautions that the test results should be interpreted "by a professional who is sophisticated in both psychology and education" (Brief Form Manual, p.6; Comprehensive Form Manual, p.5). Also, the K-TEA and K-ABC have some overlap in Mathematics and Reading, so both tests should not be administered in these areas.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously. Scores may not be reported for bilingual and/or LEP students.

References

Manual-Brief Form
Manual-Comprehensive Form

Publisher

American Guidance Service

Title	<u>Life Skills</u>	Grade	<u>9th - 12th</u>
Author	<u>Riverside Publishing Co.</u>	Time	<u>Untimed</u> <u>(40 minutes per session)</u>
Copyright Date	<u>1980</u>	Type	<u>CRT</u>

Intended Purpose of Test

The Life Skills test is designed to measure "life skills" or functional competencies in reading and math.

Description

Life Skills comes in two forms, each containing two parts which consist of 48 reading and 50 math multiple-choice questions. Both parts are included in the student test booklet. The questions relate to a variety of illustrations such as labels, ads, tax forms, and graphs which are encountered in everyday life.

The Reading part has the following objectives:

- | | |
|-------------------------|---|
| 1. Following Directions | Interpret caution, labels, directions, and instructions on household items and follow directions to use a telephone |
| 2. Locating References | Locate information in a phone book, catalog, and manual |
| 3. Gaining Information | Interpret an ad, public notice, time card, bill, lease |
| 4. Understanding Forms | Interpret a purchase agreement and tax form |

The Mathematics part has the following objectives:

- | | |
|---|--|
| 1. Computing Consumer Information | Calculate rent, income, mileage, add/subtract money, and determine "best buys" |
| 2. Applying Principles of Percentages, Interest, and Fractional Parts | Calculate time, loan interest, and sales tax |
| 3. Identifying, Estimating, and Converting Time, Currency, and Measurements | Recognize currency, determine elapsed time, calculate and convert measurements |
| 4. Interpreting Graphs, Charts, and Statistics | Interpret graphs and charts in newspapers and calculate temperature change |

Test Administration

Life Skills may be administered individually or to a group by a trained professional who is familiar with test procedures. No special materials are required.

Results may be hand-or machined-scored and are reported as raw scores and percentages. Optional score reports also include raw score frequency distributions for each school compared with districtwide percentile ranks.

Technical Information

Standardization was based on a national sample of students, grades 9 through 12, in the Fall of 1978. Groups of inmates from penal institutions and persons enrolled in adult education courses were included in the sample.

Alternate form reliability, reported for math, reading, and total score, was .79, .77, and .83, respectively. Reading and Mathematics tests were moderately correlated; .74 on Form 1 and .70 on Form 2. Alternate form reliabilities of individual objectives were low, ranging from .01 to .36 for Reading and .02 to .50 for Math. The manual stated that "scores by objective should not be considered sufficiently reliable to form a basis for making decisions about each student" (Manual, p. 21).

No validity data were presented.

Effective Use/Comments

Life Skills is designed to measure functional competencies of students in reading and math which, with other measures, may be useful for making instructional planning decisions and developing long- and shortterm objectives on the IEP in the areas of reading and/or mathematics. Information is insufficient for making performance level and evaluation of student progress decisions because of limited technical information.

Due to item content, Life Skills is not recommended for use with linguistically and culturally diverse students and LEP students.

Reference

Manual

Publisher

Riverside Publishing Co.

Title	<u>Peabody Individual Achievement Test (PIAT)</u>	Grade	<u>K-High School</u>
Author	<u>M. Dunn & C. Markwardt</u>	Time	<u>Untimed</u>
Copyright Date	<u>1970</u>	Type	<u>NRT</u>

Intended Purpose of Test

The PIAT is designed to provide a wide-range screening measure of achievement in the areas of mathematics, reading, spelling, and general information.

Description

The test is individually administered and consists of two volumes of test plates that contain demonstration, training, and test items. The test contains the following five subtests: Mathematics (84 multiple-choice items, ranging from matching numerals to identifying geometric concepts); Reading Recognition (84 items, ranging from matching letters to recognizing words); Reading Comprehension (66 multiple-choice items which require selecting a picture that corresponds to sentence meaning); Spelling (84 items which range from matching sounds with pictures to identifying correct spelling); General Information (84 orally presented questions on science, sports, arts, and social studies).

Norm Data

The standardization sample was a national sample of 2,889 male (50%) and female (50%) students attending regular classrooms in public schools and living in the nine geographic divisions which conformed to the 1967 U.S. Bureau of Census data; approximately 11.3% of the sample were black. There were approximately 200 students for each of the thirteen grade levels, from kindergarten through grade 12.

Reliability

Median test-retest reliabilities, with an interval of one month were .89 for Total Test, .74 for Mathematics, .89 for Reading Recognition, .64 for Reading Comprehension, .65 for Spelling, and .76 for General Information. The median standard error of measurement for the Total Test was 12.05; for the five subtests, the median standard errors of measurement ranged from 3.06 to 6.51. Satisfactory test-retest and split-half reliabilities for the scale also have been reported in other studies with samples of Anglo, Mexican-American, and handicapped students. No data concerning internal consistency were provided.

Validity

Content validity was based on expert opinion. Concurrent validity is yet to be demonstrated, since correlations reflected correspondence between the PIAT and the WRAT.

Special Alerts/Comments

The PIAT provides a wide-range screening measure of achievement in the areas of mathematics, reading, spelling, and general information. It serves to locate the student's general level of achievement after which a longer, more comprehensive achievement test is needed. The test findings are insufficient for making individual instructional planning decisions.

The high concentration of items suitable for lower grades adds to the discrimination power of the PIAT in more accurately determining academic achievement of educable mentally retarded students. Also, the multiple-choice pointing format on three of the five subtests makes it possible to obtain an achievement measure for students who otherwise would be difficult to assess (e.g., severe motoric problem, limited vocal ability).

Although the item content for each of the five subtests seems appropriate and although there are advantages to having an individual achievement test, the PIAT suffers in comparison to its group competitors because reliability coefficients are somewhat low. More confidence can be placed in the total score than in the subtest scores. The relationship between the content of the PIAT and the New York City curricula as reflected in the MTEs is low.

Further, because of the test's standardization and norming procedures, its validity for use with LEP students and linguistically and culturally diverse students may be inappropriate. Scores may not be reported for bilingual and/or LEP students.

References

Manual
Salvia & Ysseldyke, 1985
Sattler, 1982

Publisher

American Guidance Service, Inc.

Title	<u>Stanford Achievement Test (SAT)</u>	Grade	<u>1st - 9th</u>
Author	<u>E. F. Gardner, H. C. Rudman,</u> <u>B. Karlson, J. C. Merwin,</u> <u>R. Callis, & C. S. Collins</u>	Time	<u>Varies according to</u> <u>Level</u>
Copyright Date	<u>1982</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The SAT is an achievement test providing information on reading, mathematics, spelling, science, social studies, and listening comprehension.

Description

The SAT is a group administered test consisting of multiple choice questions. A writing assessment is also available. Separate booklets are provided which include all reading tests, all mathematics tests and the listening tests. The Basic Battery (i.e., all the subjects except Science and Social Science) takes from two to four hours. The complete battery can take five hours.

Norm Data

The standardization group consisted of over 450,000 students. This norm sample took into account geographic region, public and private school system, major ethnic groups and school district size.

Separate norms were established for Fall, Mid-Year, and Spring testing at grades Kindergarten through 12.

Reliability

Reliability for the SAT was established by computing KR-20 internal consistency coefficients and alternate form reliability coefficients for each test level. These coefficients ranged from .76 to .96.

Validity

Content validity was established by submission of test items to a panel of subject-matter experts. The test items were also submitted to minority group members who judged the items for appropriateness of content for different cultural groups.

Correlation with previous SATs and current and previous Metropolitan Achievement Tests was used to establish empirical validity.

Special Alerts/Comments

The SAT provides for a comprehensive evaluation of skill development in key content areas which is sufficient for screening and evaluation of student progress decisions.

There is a high degree of correspondence between test items and the MTEs.

Given its group administration, the information obtained from the SAT is insufficient for decisions regarding individual instructional planning and determination of performance levels for students who may have handicapping conditions. For these students, further individual assessment is necessary.

Special editions of the SAT are available for the blind or partially sighted and for the deaf. The SAT has also been normed on these populations.

Although the norms are not representative of students with handicapping conditions, the scores indicate the achievement of a student in relation to other students of similar age/grade.

Caution should be exercised when early childhood students and recently arrived LEP students are included in the testing population, since the degree of exposure to English language skills influences performance. Scores may not be reported for bilingual and/or LEP students.

References

Manual
Buros, 1985
Salvia & Ysseldyke, 1985

Publisher

The Psychological Corp.

READINESS ASSESSMENT

Readiness assessment measures a student's ability to function in a school environment and begin academic instruction. Most instruments included in this section are multiskill batteries.

Title Diagnostic Inventory of Basic Skills (DIBS)

Grade K-6th

Author A.H. Brigance

Time Untimed

Copyright Date 1977

Type CRT

Intended Purpose of Test

The DIBS is designed to assess basic readiness and academic skills of students from kindergarten to sixth grade. It was designed to help teachers determine the student's level of achievement and to maintain a continuous progress record for each student.

Description

The DIBS consists of a test manual, student record books, and a class record book.

The test manual is published in a sturdy flip-over binder so that the examiner and student may be seated opposite each other during test administration. Each skill area is arranged in a developmental and sequential hierarchy and includes the following:

Readiness	color recognition, visual memory, body image, gross motor coordination, identification of body parts, directional/positional skills, fine motor skills, verbal fluency, verbal directions, articulation of sounds, personal data response, sentence memory, counting, alphabet, numeral recognition, number comprehension, recognition of lower case letters and upper case letters from dictation
Reading	word recognition, oral reading comprehension, word analysis, vocabulary
Language Arts	handwriting, grammar mechanics, spelling, reference skills
Math	number concepts, operations, measurement (money, time, calendar, and linear, liquid, and weight measurement), geometry

Test Administration

The Student Record Book lists each individual skill with a grade level notation indicating the grade at which the skill is normally taught. An additional guide, Skills by Curriculum Level, is intended to help the

assessor determine where to begin testing. The examiner's page includes the skill, performance objective, directions, scoring, timing, and mastery levels for each skill. Although the test was designed for one-to-one administration, many of the subtests may be group administered with written rather than oral responses. Testing should begin at the anticipated level of 90% success and continue until the ceiling level (no success) and the basal level (100% success) are obtained. A grade level profile for each student and a Class Record Book enables the teacher to plan instruction and plot the student/class progress. No special training or special material is required to administer the test.

Technical Information

The DIBS was developed in Northern California to meet the needs of school programs requiring diagnosis, record keeping, and individualized instruction. Three unpublished versions were piloted with elementary school students in Title I, California Early Childhood Education (K-3), Indian Education, and Special Education programs between 1970-1975. Content validity, developed during field test revisions, was based upon feedback from 66 teachers and other curriculum personnel from three schools in Northern California, an unspecified number of school personnel in the Boston, Massachusetts area, and Monitor and Review Teams from the California State Department of Education. Reliability was not addressed. Information on how mastery criteria were obtained was also omitted.

Effective Use/Comments

The information provided by DIBS is sufficient for developing short- and longterm objectives on the IEP in the areas of readiness, reading, language arts, and mathematics. The information is not sufficient for making screening and performance level decisions. This test follows the sequence of the MTEs. While instructional objectives are specified, the manual does not provide strategies for classroom instruction.

The DIBS is not as comprehensive as later inventories developed by Brigance. It provides classroom teachers of grades K-6, however, with an effective instructional management system in the basic skills, as the title suggests.

Some of the readiness and math subtests may be given to linguistically and culturally diverse students and LEP students when individually administered. Other subtests for these populations must be carefully selected and test results interpreted with caution.

Reference

Manual

Publisher

Curriculum Associates, Inc.

Title	<u>Inventory of Early Development (IED)</u>	Age	<u>Birth to 7 Years</u>
Author	<u>A.H. Brigance</u>	Time	<u>Untimed</u>
Copyright Date	<u>1978</u>	Type	<u>CRT</u>

Intended Purpose of Test

The IED is designed to be used with infants and children below the developmental level of seven years to assess psychological and early academic skill development. The results provide information to identify, assess, and develop individualized educational programs for students with learning handicaps.

Description

The IED is assembled in a sturdy looseleaf binder with convenient flip over pages for one-to-one administration. It includes 98 developmental skill sequences within the following categories:

Psychomotor	pre-ambulatory motor skills and behaviors, gross motor skills and behaviors, fine motor skills and behaviors, self-help skills
Language	pre-speech, speech and language skills, general knowledge and comprehension
Early Academic Skills	readiness, basic reading skills, manuscript writing, basic math

The manual and the developmental record book indicate the age level that each sub-skill is usually acquired.

Test Administration

The IED is too comprehensive to be administered in its entirety. The examiner determines the skill sequences and skill levels depending upon the purposes for assessment, age, handicapping conditions, and informal observational data about the student. The manual provides background information and suggestions to follow regarding other methods of assessment for each skill category: parent interviews, observations, and specific formal and informal procedures. Although the author stated that the IED does not require specialized training in testing, the assessment should be conducted under trained professional supervision. Specialized materials are not required. Some of the assessments may be adapted for group administration. No alternate forms are available.

Recording data in the developmental record book using a color coding system provides an on-going, easy to interpret graphic record of progress for each student.

Technical Information

The same design was followed in developing the IED as other Brigance inventories. Extensive research was conducted in order to establish and validate the skill sequences and the developmental ages. Content validity was further developed through field test revisions based upon feedback from professionals representing different disciplines and different types of programs and services for children below the 7-year-old developmental age in 16 states.

Reliability was not addressed.

Mastery levels depend upon the skill that is assessed. It is 90% in arithmetic, 97% in reading, 100% in most other skills. In some skills the mastery decision is left to the examiner's discretion.

Effective Use/Comments

The information provided by the IED is sufficient for developing short-term objectives on the IEP in the area of psychomotor, language, and early academic skills. The information provided by the IED is not sufficient for making individual screening performance level, instructional planning, and evaluation of progress decisions.

This test follows the sequence of the MTEs for kindergarten and grade 1. While objectives are specified, instructional strategies are not provided.

Although the IED may be administered to linguistically and culturally diverse students and LEP students preferably by an examiner who is fluent in the student's primary language and/or understands the student's culture, results should be interpreted with caution.

Reference

Manual

Publisher

Curriculum Associates

Title	<u>K & 1 Screen (Screen)</u>	GRADE	<u>Pre K-1st</u>
Author	<u>A.H. Brigance</u>	Time	<u>10-20 Minutes</u>
Copyright Date	<u>1982</u>	Type	<u>CRT</u>

Intended Purpose of Test

Screen was designed to assist a screening team in the acquisition and evaluation of data necessary for the admission of a student to kindergarten or first grade. It provides an overall picture of the student's development in the following areas: language development, motor ability, number skills, body awareness, and auditory and visual discrimination.

Description

The manual is published in a sturdy flip-over binder so that the examiner and student are seated opposite each other for one-to-one administration. It includes the following components:

Basic Screening Assessment

Personal data response	Color recognition
Picture vocabulary	Visual discrimination
Visual-motor skills	Gross motor skills
Draw a person (body image)	Rote counting
Identification of body parts	Recitation of alphabet
Following verbal directions	Numeral comprehension
Recognition of lower case and upper case letters	Auditory discrimination
Syntax and fluency	Printing personal data
	Numerals in sequence

Advanced Assessments (Optional use for advanced students)

Response to picture	Articulation of sounds
Basic preprimer vocabulary	Oral reading at preprimer/primer level
Basic number skills	

The skills assessed are basically the same as those covered by IED, DIBS and CIBS. Screen does not screen for physiological development (physical examination, vision, hearing, speech impairment), which requires appropriately trained personnel. No alternate forms are available.

Test Administration

Screen should be scheduled for administration at the time of year that best serves the needs and purposes of the school program. It is not recommended that Screen be routinely administered to all students as a posttest. If the teacher/examiner deems that re-screening is necessary, the author suggested a waiting period of six weeks. Specialized training in test administration is an asset but not considered a prerequisite.

Training and supervision should be given, if a paraprofessional is assisting.

One-to-one testing is preferred, although specific tests, which are identified in the manual, may be group administered and/or assessed as part of the daily instructional program. No specialized material or equipment is needed. The author offered a strategy, "The Station Method", for screening large numbers of students.

Point values are assigned to each skill to yield a possible total score of 100 so that students within a given group may be ranked. The author recommended that students who score below 60 be evaluated in more detail. He cautioned, however, that referrals for additional assessment should not be based solely upon the screening score. Age, maturity, and observations by the assessor, teachers, and parents should be considered.

Appropriate parent and teacher checklists are furnished and all significant testing data are recorded on the Student Data Sheet by the assessor. Charts are provided in the manual showing the correlation between the skills assessed by Screen and the IED and IBS so that the examiner may refer to additional tests, if needed.

Technical Information

Screen was developed as a result of numerous requests received by users of the IED and the DIBS to assemble one comprehensive inventory appropriate for kindergarten and first grade screening at the time that state educational departments were developing plans to comply with PL 94-142's "Search and Serve" mandate. The screening procedures were piloted at two elementary schools in Holland, Pennsylvania and Harlan, Kentucky. Content validity was further developed during the field test revisions based upon feedback from professionals in 45 regular and special education programs in 14 states. Reliability was not addressed.

Effective Use/Comments

Lack of technical information precludes the use of Screen for making screening, performance level, instructional planning, and evaluation of progress decisions.— The information provided by Screen is sufficient for developing short-term objectives on the IEP in the area of readiness.

The sequence of the MTE's is closely followed. The manual does not include strategies for instruction.

Use of Screen with linguistically and culturally diverse students and LEP students should be undertaken by personnel who are fluent in the student's dominant or native language. Caution should be exercised in interpreting results of these students.

Reference

Manual

Publisher

Curriculum Associates

Title	<u>Pediatric Examination of Education Readiness(PEER)</u>	Ages	<u>4 - 6 years</u>
Author	<u>M.D.Levine & E.A Schneider</u>	Time	<u>10-20 Minutes</u>
Copyright Date	<u>1982</u>	Type	<u>CRT</u>

Intended Purpose of Test

The PEER is designed to identify specific services needed by children between four and six years of age by integrating medical, developmental, and neurological findings with observations of behavioral adjustment and behavioral style. The authors indicated that PEER may be used for diagnostic evaluation, prediction, screening, and training.

Description

The PEER is a multidimensional examination consisting of six broad developmental areas with a total of 29 test items. They are:

<u>Category</u>	<u>Item</u>
Orientation	identify body part; imitative finger movement; graphesthesia; stereognosis; visual tracking
Gross Motor	walk on heels and toes; stand on one foot; pronation-supination; motor stance; catch ball
Visual-Fine Motor	visual matching; manipulate sticks; copy figures; draw from memory; block construction
Sequential	finger opposition; object span; word span
Linguistic	spatial directions; name pictures; complex sentences; categorize; temporal directions
Preacademic Learning	name symbols; name days of the week; count fingers; count aloud; write letters

Developmental attainment is determined on the basis of performance in each category. Three levels are identified: definite concern; possible concern; no concern. Categories considered as strengths are also identified.

In addition to including developmental attainment for the six categories, the examiner rates the subject on selective attention, processing efficiency, and adaptation. Results of a physical examination including vision, auditory, and neurological testing are reported at three levels: prominent findings, moderate findings, and few/no findings.

The examiner also determines the child's ability in input, storage, and output processes through task analysis scoring.

Test Administration

The PEER was designed to be administered by a physician or nurse. However, psychologists and special educators may administer portions other than the physical and neurological examinations. Instructions for administration and scoring test items are clear. The record form for reporting scores is well organized. However, organization of test administration and scoring in the manual may cause difficulties to using the material during testing.

Technical Data

The PEER was field tested among 386 children in Brookline, Massachusetts prior to kindergarten. Reliability and validity of adequate magnitudes were determined by developing an aggregate score of the number of areas of "concern". Mean interrater reliability of areas of concern was .89. Predictive validity coefficients of the number of areas of concern and the Kindergarten Performance Profile ranged from .34 to .57.

The method used for determining a definite "concern", "possible concern", or "no concern" and "strength" was not addressed. Similarly, the basis for the maximum rating for interpreting task analysis scores was omitted.

Effective Use/Comments

The PEER provides information about a young child which may be useful for supplementary, descriptive purposes. The information is insufficient for making screening, performance level, instructional planning, and evaluation of progress decisions. Results may be used for writing shortterm objectives on the IEP in the area of readiness.

The PEER is based on a medical model. The authors cautioned users on interpreting neurological data.

Use of the PEER with culturally and linguistically diverse children and LEP children has not been demonstrated.

Reference

Manual

Publisher

Educators Publishing Service

READING

Many different skills comprise the complex behavior defined as "reading". Different aspects of reading are measured by various tests. Areas assessed include comprehension, word-attack or analysis, syllabication, decoding, and reading rate. Some decoding tests assess a student's ability in phonics.

Title	<u>Bieger Test of Visual Discrimination (BTVD)</u>	Age/Grade	<u>Any non or poor reader</u>
Author	<u>E. Bieger</u>	Time	<u>Untimed</u>
Copyright Date	<u>1982</u>	Type	<u>CRT</u>

Intended Purpose of Test

The BTVD is designed to gather information about the visual discrimination ability of poor or nonreaders through an investigation of letters and words.

Description

The BTVD consists of seven multiple-choice tests, three discriminating between letters, four discriminating between words. Test variations are based on the contrasts between choices and the stimulus letter or word.

- | | |
|--------|---|
| Test 1 | 26 lower case letters differentiated from choices on the basis of large contrasts; e.g., o: l v o z or t: c y s t |
| Test 2 | 17 lower case letters differentiated from choices on the basis of small contrasts; e.g., t: f x t or o: c a o |
| Test 3 | 10 lower case letters which are almost identical; e.g., b: p q b d or b: d b p g |
| Test 4 | 13 2-to 4-letter words differentiated from choices on the basis of large contrasts; e.g., door: lift eves like door |
| Test 5 | 12 2-to 4-letter words differentiated from choices on the basis of small contrasts; e.g., wet: wet vex met |
| Test 6 | 11 3-letter words which are almost identical to choices in configuration; e.g., gap: gag gap gap gad |
| Test 7 | 23 2-to 5-letter words differentiated from choices on the basis of sequence; e.g., two: out wet two tow |

Test Administration

The BTVD may be given to small group of students as well as individually. Mastery and criterion scores have been established for each subtest. If a student does not meet the criterion score on a subtest, no further testing should occur.

Mastery is considered 100%. Students who achieve between the criterion score and mastery are deemed to have an "adequate" skill level, but are eligible for remediation in the skill.

Technical Information

The test is based upon research and theory in perceptual learning, thereby establishing construct validity. The inclusion of operational definitions of the processes measured on the test also supports its construct validity. Content validity is established through the representativeness of the content of the test.

Consistency of test scores was investigated through calculating the percentages of students whose scores differed within one week of re-testing. Most differences were less than six percent.

Criteria for establishing mastery, adequacy, and inadequate or poor performance were not addressed.

Effective Use/Comments

The information provided in the BTVD is sufficient for making instructional planning decisions in the area of discrimination of letters and words. The information is useful for developing shortterm objectives on the IEP in this area. The information is not sufficient for making performance level decisions in this area.

In addition to specifying objectives, the manual provides specific strategies regarding instruction.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

Reference

Manual

Publisher

Stoelting Co.

Title	<u>Diagnostic Reading Scales (DRS)</u> <u>(Revised Edition)</u>	Grade	<u>1-7</u>
Author	<u>G.B. Spache</u>	Time	<u>Untimed (full test</u> <u>1 hr. approx)</u>
Copyright Date	<u>1981</u>	Type	<u>NRT</u>

Intended Purpose of Test

The DRS is designed to assess oral and silent reading skills and auditory comprehension. The test provides an estimate of the student's instructional, independent, and potential reading levels.

Description

Individually administered, the DRS consists of three word recognition lists, two sets of reading selections (two in each set), and twelve word analysis and phonic tests. The examiner uses an Examiner's Record Book and Manual; the student uses a Student's Reading Book. Administration of the DRS is straightforward; scoring is uncomplicated and can be done during testing.

Norm Data

Norm data were based on the 1972 edition of the DRS. The standardization population consisted of 2,081 students in rural and urban schools in Florida, Rhode Island, Georgia, and New York. Testing was done by experienced reading clinicians. Norms apply only to the word recognition lists and reading passages which yield grade levels from raw scores.

The nature of the population of the 1981 edition was not fully described.

Reliability

For the earlier edition, internal consistency was reported as .87 to .96 on the word recognition lists and as .98 to .99 on the reading passages. Test-retest reliability after a 10-week interval was .84 to .88.

Validity

For the earlier edition, concurrent validity coefficients were reported as follows: .65 to .78 with the Durrell word list as per teacher judgement, .92 to .93 with the Durrell Listening-Reading Test, and .90 to .92 with the Gates-McKillop Reading Diagnostic Test. Content validity was demonstrated for this edition through reports of test development procedures.

Special Alerts/Comments

The DRS provides an estimate of the student's instructional, independent, and potential reading levels through an assessment of oral and silent reading and auditory comprehension skills. Because of the technical inadequacies in the norming procedures and incomplete data on reliability and validity, the data are insufficient for determining performance levels in reading. The DRS may be helpful in assisting in instructional planning decisions as an informal tool.

The CRS give a comprehensive overview of reading skills highlighting a student's strengths and weaknesses. Selected subtests of the DRS may be administered and scored quickly and yet give a rather accurate picture of a student's reading abilities. It is also a lightweight, portable test which can be re-used. The examiner should be trained in the teaching of reading to analyze information obtained through the CRS. As an informal tool, certain subtests may be useful with linguistically and culturally diverse students whose development of English language skills has been determined.

References

Manual
Salvia & Ysseldyke, 1985

Publisher

CTB/McGraw Hill

Title	<u>Durrell Analysis of Reading Difficulty</u> <u>(DARD) - 3rd Edition</u>	Grade	<u>1st - 6th</u>
Author	<u>D.O. Durrell & J.H. Catterson</u>	Time	<u>Timed and Untimed</u> <u>(30-90 minutes)</u>
Copyright Date	<u>1980</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The DARD is designed to estimate the general level of reading achievement and to identify strengths and weaknesses for the purpose of planning a remedial reading program.

Description

The DARD is individually administered. No specific qualifications for the examiner are indicated. An individual test booklet is provided for summarizing the results. The booklet includes a profile chart, a checklist of instructional needs, and a general history form.

Test directions and procedures are not difficult but the examiner must familiarize himself/herself with the reading content and suggested marking system.

As a criterion referenced test, the DARD provides checklists for recording observations of reading difficulties and evaluations of the following reading abilities: Oral Reading, Silent Reading, Listening Comprehension, Word Recognition/Word Analysis, Listening Vocabulary, Pronunciation of Word Elements, Spelling, Visual Memory of Words, Auditory Analysis of Words and Word Elements and Prereading Phonics Abilities: Syntax Matching, Identifying Letter Names in Spoken Words, Identifying Phonemes in Spoken Words, Naming Lower Case Letters, and Writing Letters from Dictation.

As a norm referenced test, the DARD provides an estimate of the student's reading level.

Norm Data

More than 1,200 students, approximately 200 per grade, in six communities in six states located throughout the U.S., were tested in the standardization process. Factors such as language background, socioeconomic status, ethnic characteristics, and curriculum emphasis were taken into consideration. The nature of the population was not detailed.

Reliability

While in previous editions no reliability data were reported, the latest manual reported internal consistency coefficients ranging from .63 to .97 for individual subtests; among thirteen subtests, correlations of five were about .85.

Validity

Content validity was indicated through test development procedures. Predictive validity for selected prereading subtests ranged between .55 and .65. Using the Metropolitan Reading Test as the criterion measure, concurrent validity was indicated for primary tests and intermediate tests (r 's ranged between .22 and .94 and between .42 and .96, respectively).

Special Alerts/Comments

According to the manual, the DARD is best used as a criterion referenced test. Given its technical limitations as a norm referenced test, the DARD may be used to provide qualitative information about a student's reading strengths and difficulties. While it is easy to administer, interpretations are best done by an experienced teacher of reading who can provide directions for instructional planning.

The information provided in the DARD is sufficient for making instructional planning decisions and developing short- and longterm objectives on the IEP.

As an informal tool, certain subtests may be useful with LEP students and linguistically and culturally diverse students whose development of English language skills has been determined. Caution should be used in interpreting results because of cultural loading of reading passages. Scores may not be reported for bilingual and/or LEP students.

References

Manual
Salvia & Ysseldyke, 1985

Publisher

The Psychological Corp.

Title	<u>Gates-MacGinitie Reading Tests</u> <u>(second edition)</u>	Grade	<u>K-12th</u>
Author	<u>H. MacGinitie</u>	Time	<u>55-65 minutes</u> <u>per level</u>
Copyright Date	<u>1978</u>	Type	<u>NRT</u>

Intended Purpose of Test

The purpose of this test is to assess the student's general level of reading skills in vocabulary and comprehension.

Description

The test has seven levels ranging from kindergarten to 12th grade with one "out of level" test. With the exception of Basic R (the earliest level), each level consists of two subtests, vocabulary and comprehension. Basic R includes, in addition to vocabulary and comprehension, letter recognition and letter sounds. The tests are generally administered as group tests, responded to by the students either in booklets or on answer sheets.

Norm Data

This edition of the test was standardized from October 1976 to May 1977 at varying times for each of the levels. The total number of students in each standardization testing was approximately 65,000. The norming samples were obtained from a stratified sampling design based primarily on the Fourth Count of the 1970 U.S. Census, which provided data on the basis of school district boundaries. The sampling plan stratified the districts on four variables: geographic location, enrollment size, family income, and years of schooling completed by the adult population. Black and Hispanic students were included in proportion to their representation in the population. Eighty-six school districts participated in the norming.

Reliability

Reliability coefficients (KR20) were computed for each level of the test. They ranged from .88 to .95. Alternate form reliability coefficients for the total test at all levels were above .89. Test-retest reliability ranged from .77 to .89.

Validity

With respect to validity, the authors stated: "To the extent that the test measures what the user wants measured and nothing else, the test is valid. A reading test is given in order to evaluate how well students in a particular school are learning to read. It is valid for that school to the school's goals in teaching reading" (Manual).

Special Alerts/Comments

The results from the Gates-MacGinitie Reading Tests may be useful for screening and evaluation of student progress decisions with respect to reading skills in vocabulary and comprehension. The test findings are insufficient for making individual instruction planning decisions.

There is a high degree of correspondence between test items and the MTEs. Although the norms are not representative of students with handicapping conditions, the scores indicate the achievement of a student in relation to other students of similar age/grade.

Caution should be exercised when early childhood students and recently arrived LEP students are included in the testing population, since the degree of exposure to English language skills influences performance. Scores may not be reported for bilingual and/or LEP students.

References

Manual
Salvia & Ysseloyke, 1985

Publisher

Riverside Publishing Co.

Title	<u>Gates-McKillop-Horowitz Reading Diagnostic Tests (2nd edition)</u>	Grade	<u>2-6th</u>
Author	<u>A. Gates, A. McKillop, & E.C. Horowitz</u>	Time	<u>timed and untimed (approx. 30-60 mins)</u>
Copyright Date	<u>1981</u>	Type	<u>NRT</u>

Intended Purpose of Test

The Gates-McKillop-Horowitz is designed to assess skill development in reading.

Description

The tests consist of 14 individually administered subtests which tap the following reading skills: oral reading, word attack skills, auditory blending, and word equivalent sounds. Two forms of the test, I and II, are available. Depending on student age and level of skill development, the examiner selects particular subtests to give each student.

Norm Data

In 1979, 600 students in grades 1 through 6, a majority of whom attended urban, private schools, were tested. Approximately half were male. About one-third were black.

Reliability

Test-retest reliability for a small sample of students on the Oral Reading Test was reported as .94. Based upon 50 Oral Reading protocols, interrater reliability was reported as 94% agreement between two raters. However, Salvia and Ysseldyke (1985) have questioned the reliability of this test.

Validity

Comparison with Gates-MacGinitie and Metropolitan Reading Test Scores indicated correlations of .68 to .98 at the various grade levels; correlations were particularly high at the lower grades. However, the validity of this test for older or better readers is questionable.

Special Alerts/Comments

The information obtained from the Gates-McKillop-Horowitz may be useful for making instructional planning and evaluation of student progress decisions with respect to the reading skills measured. Given the technical inadequacies, the data are insufficient for making performance level decisions. While administration is relatively easy, interpretation and scoring are complex and difficult, requiring expertise in reading diagnosis.

As an informal tool, certain subtests may be useful with LEP students and linguistically and culturally diverse students whose development of English language skills has been determined. These comments on test use are based upon the 2nd edition of this test only. Scores may not be reported for bilingual and/or LEP students.

References

Manual
Salvia & Ysseldyke, 1985
Mitchell, Jr., 1985

Publisher

Harper and Row

Title Gilmore Oral Reading Test

Grade 1st - 8th

Author V. Gilmore & E.C. Gilmore

Time 15 - 20 minutes

Copyright Date 1968

Type NRT

Intended Purpose of Test

The Gilmore is designed to assess skill development in oral reading. It provides measures of oral reading accuracy, comprehension, and rate.

Description

The Gilmore has two equivalent forms, C and D. Each form is composed of ten oral reading paragraphs which increase in difficulty. These ten paragraphs form a continuous story. After each paragraph, there are five questions which measure the student's comprehension of the material. There is an individual record blank which is used to measure the student's time for each paragraph, the errors made in reading, and comprehension of the material read.

Norm Data

Standardization of the Gilmore was conducted on 4,455 students from various socioeconomic backgrounds; they attended 18 schools located in six school systems. Approximately half the students were given Form C, while the other half were given Form D. No data on sex, ethnic/racial background, or reading curriculum of the students in the norm sample were given in the test manual.

Reliability

Alternate form reliability, based on samples of 51 students in the third grade and 55 students in the sixth grade, ranged from .53 to .94. No information was given regarding the test-retest reliability or internal consistency.

Validity

The manual cited validity data based upon comparisons of student performance on an early edition of the Gilmore with the Gray Oral and the Durrell Analysis of Reading Difficulty. No validity data were provided for the 1968 edition.

Special Alerts/Comments

Given the technical inadequacies of the Gilmore, it is insufficient for determining a performance level in reading. However, the results of the Gilmore may be useful for providing supplementary, descriptive information about the kind of oral reading errors made by the student.

Salvia and Ysseldyke (1985) indicated the Gilmore "may provide the experienced examiner with diagnostic information with which to construct instructional hypotheses" (p. 349).

Since the research on the relationship between oral and silent reading has yielded contradictory results in terms of whether comprehension in one is greater than, equal to, or less than the other, it is recommended that if an individual performs poorly on the Gilmore, he or she should be given a silent reading test. If the results of this test indicate that silent reading ability is what would be expected for his/her age and ability, further standardized assessment of reading is probably warranted, even if oral reading ability is below silent reading ability.

As an informal tool of oral reading, the Gilmore may be useful in assessing LEP students and linguistically and culturally diverse students whose development of oral English skills is known. Caution should be used in interpreting results because of cultural loading of reading passages. Scores may not be reported for bilingual and/or LEP students.

References

Manual
Salvia & Ysseldyke, 1985

Publisher

The Psychological Corp.

Title	<u>Gray Oral Reading Test-Revised (GORT-R)</u>	Age	<u>6 1/2 to 18 years</u>
Author	<u>J. Wiederholt & B.R. Bryant</u>	Time	<u>15-30 minutes</u>
Copyright Date	<u>1986</u>	Type	<u>NRT</u>

Intended Purpose of Test

The GORT-R is designed to assess oral reading speed and accuracy, oral reading comprehension, total reading ability, and oral reading miscues.

Description

The GORT-R consists of 13 reading passages of increasing difficulty. Two alternate equivalent forms (A and B) are available. Five comprehension questions follow each passage. The examiner should be thoroughly familiar with the test and the manual. Any prompting of the student must be avoided, while encouragement and praise can be offered.

Basal and ceiling levels are attained in order to shorten testing time. Each student's oral reading performance is recorded and summarized in a Profile/Examiner Record Form booklet. The Record Form provides space for pertinent information relating to the student, such as GORT-R scores, other test scores, summaries of the student's miscues, comments and recommendations. A student's oral reading performance can be reported in terms of raw scores, percentiles, standard scores, and quotients.

Norm Data

Norm data for the GORT-R are much more extensive than for its predecessor. Over 1400 students from 7 to 17 years of age were included in a sample, representative of the four major geographic region of the U.S. Age, race, ethnicity, sex and type of residence (i.e. urban vs rural) were reported.

Reliability

Internal consistency coefficients using Guilford's formula ranged from .92 to .97 for the total test. Alternate-form reliability coefficients were significant at the .001 level, providing adequate evidence that the GORT-R forms are equivalent.

Validity

The content validity of the GORT-R was established by systematic and controlled item and passage selection. Concurrent validity coefficients were determined through comparisons of the GORT-R with the Iowa Tests of Educational Development, the Formal Reading Inventory, and the Comprehensive Test of Basic Skills. The correlations provide adequate evidence of concurrent validity. The construct validity of the GORT-R has been established by comparisons of the GORT-R to age, language ability, total school achievement, and intelligence.

Special Alerts/Comments

The GORT-R is intended to help identify students who are deficient in oral reading skills. The test can be utilized to identify strengths and weaknesses and to assess student progress in reading.

Since the research on the relationship between oral and silent reading has yielded contradictory results in terms of whether comprehension in one is greater than, equal to, or less than the other, it is recommended that if an individual performs poorly on the GORT-R, he or she should be given a silent reading test. If the results of this test indicate that silent reading ability is what would be expected for his/her age and ability, further standardized assessment of reading is probably not warranted, even if oral reading ability is below silent reading ability.

As an informal tool of oral reading, the GORT-R may be useful in assessing LEP students and linguistically and culturally diverse students whose development of oral English skills is known.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously. Scores should not be reported for bilingual and/or LEP students.

Reference

Manual

Publisher

PRO-ED

Title Individualized Criterion Referenced
Test: Reading (ICRT)

Grade K - 8th

Author Educational Development Corp.

Time Untimed

Copyright Date 1980

Type CRT

Intended Purpose of Test

The ICRT is designed for reading diagnosis and prescription. It also enables a teacher to target appropriate instruction toward objectives that students have not mastered and assists in grouping students for instruction on the basis of specific need rather than on general level of achievement.

Description

The ICRT consists of 38 tests measuring 345 objectives in levels 1 through 8. Tests are grouped in booklets consisting of 16 items for determining mastery of eight objectives. Two forms of each booklet are available.

Reading skills are organized into four strands as indicated below:

Phonetic Analysis	Initial consonants Final consonants Regular vowels Irregular vowels Consonant blends and digraphs
Structural Analysis	Plurals, verb variants and possessives Contractions and compound words Syllabication Roots and affixes
Word Function Skills	Context clues Synonyms and antonyms
Comprehension	Reading for detail Relationship of time and space Inference Main idea Attributes Causes and effects Classification and interpretation Author's purpose Applications

Skills are referenced to a variety of texts and other instructional material. Mastery of an objective is defined as a perfect score on a 2-item test.

Test Administration

Each student is tested, individually or in a group, on five booklets in one level to obtain information on approximately 40 objectives.

The manual describes in detail the type of score reports based on computer scoring that are available. Printouts indicate objectives mastered and not mastered and three instructional references for each.

Technical Information

Items for the program were field tested on approximately 80,000 students in grades 1 to 8 in six districts in Orange County, California. According to the manual, the content validity of the continuum of test objectives depends on each school district's reading curriculum.

A 1979 study on the quality of objectives reported that more than 90% of the objectives were judged to be highly appropriate, clearly written, and correctly sequenced.

A reliability study of 400,000 students indicated that for more than 98% of the objectives, more than 80% of the students were consistently classified as "master or non-master." However, Hartley and Kling as quoted in Buross (1978, p.1265) stated that the reliability reported was inadequate and misleading. No information was given on the criteria for setting the level of mastery.

Effective Use/Comments

The ICRT provides useful information about reading skills of students in grades 1 through 8. It may be helpful for developing long- and shortterm IEP objectives. Based upon the technical information, results are not sufficient for making screening and level of performance decisions. The teacher should be selective in the choice of objectives and stringent in making judgments that mastery has been achieved.

It is not recommended for use with linguistically and culturally diverse students and LEP students.

References

Manual
Buross, 1978

Publisher

Educational Development Corp.

Title	<u>Roswell-Chall Diagnostic Reading</u>	Age/Grade	<u>1st - 4th Grade (and</u>
	<u>Reading Test (Revised & Extended)</u>		<u>for students at higher</u>
Author	<u>F.G. Roswell & J.S. Chall</u>		<u>levels where difficulty in</u>
			<u>decoding and basic word</u>
			<u>recognition is suspected)</u>
Copyright Date	<u>1978</u>	Time	<u>10 minutes</u>
		Type	<u>CRT</u>

Intended Purpose of Test

The Roswell-Chall is designed to assess decoding and word recognition skills.

Description

The test consists of 14 subtests, arranged in order of difficulty. The first 10 include decoding skills. They test word-recognition, single consonant sounds, consonant digraphs and blends, short and long vowel sounds, rule of silent e, vowel digraphs, common diphthongs and vowels controlled by e, and syllabication. The remaining four subtests evaluate the student's ability to name capital and lower case letters, encode single consonants, encode phonically regular words of three letters each.

Mastery varies for each subtest and is indicated in tables in the manual. Total decoding mastery is represented by 71 out of 82 correct responses.

The Roswell-Chall is packaged as a booklet, which includes the subtests themselves and a page for recording test results. It is accompanied by a manual of instruction for two forms (A and B).

Test Administration

The test is administered individually and requires approximately 10 minutes for administration, scoring, and interpretation. No specific materials are required, other than two copies of the test booklet and writing implements. One test booklet is used by the examiner to record the student's responses, while the other is used by the student.

It is suggested that not every subtest be given to every student. When results of earlier subtests indicate failure, it is unnecessary to administer those assessing more difficult skills. Furthermore, when the subject offers four consecutive incorrect responses within a subtest, the subtest may be discontinued.

Specific instructions for administration and scoring are presented in the manual.

Technical Information

This revised edition of the Roswell-Chall contains five subtests not included in the original. Data were collected on a sample of 203 students in Grades 1 through 4 and a clinic population of 46 in grades 2.2 to 10.2. Alternate-form reliability was .98 for the total test, while it ranged from .84 to .94 for the various subtests.

Correlation with the Metropolitan Readiness Test and the Gray Oral Reading Test was reported as .70 and .84, respectively.

Mastery for the decoding subtests as well as the word recognition subtest was determined from average scores attained by the standardization population on both forms.

Effective Use/Comments

Although the Roswell-Chall may be useful for obtaining descriptive information about the student with respect to decoding and encoding, the test findings are insufficient for making performance level and instructional planning decisions in these areas. The test measures knowledge of certain elements essential to word attack, but not the student's ability to apply these elements. It is very possible that the words used are part of the student's sight vocabulary. Perhaps application of these skills would be better tested through the use of nonsense words.

Although this test may be administered, scored, and interpreted in a short period of time and the manual of instructions is easy to follow, it would appear that the teacher would benefit from knowledge of the skills being tested since the scoring of the Roswell-Chall is subjective. The test does not follow the sequence of the MTE's.

It is suggested that this test can best be used for students making normal progress in reading and as part of an informal reading inventory. In this way, the teacher will have other measures of the student's ability, including standardized silent and oral reading test.

As an informal tool, certain subtests may be useful with LEP students and linguistically and culturally diverse students whose development of English language skills has been determined.

Reference

Manual - Technical Supplement

Publisher

Essay Press, Inc.

Title	<u>Stanford Diagnostic Reading Test (SDRT)</u>	Grade	<u>1.5 - 12</u>
Author	<u>B. Karlsen, R. Madden, & E. Gardner</u>	Time	<u>2 hours</u>
Copyright Date	<u>1978</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The SDRT is designed to provide a diagnostic reading assessment of skills in decoding, vocabulary, comprehension, and rate.

Description

The SDRT is a diagnostic tool which can be group administered. It is comprised of four overlapping levels corresponding to grade level and achievement. The SDRT is both norm referenced and criterion referenced. Students respond on test booklets. The manual at each level lists instructional objectives.

Norm Data

The test was standardized on approximately 31,000 students in a random sample of 55 schools. The sample took into account socioeconomic status, school enrollment, and geographic region.

Reliability

Two forms of reliability were reported. Internal consistency coefficients for all subtests exceeded .90 except for Auditory Vocabulary which was reported above .85. Alternate form reliability coefficients ranged from .75 to .94.

Validity

Criterion-related validity was established by comparing performance on the SDRT with the SAT reading subtest. Correlations ranged from .39 to .98. Content validity was based on the relationship to the reading curricular content.

Special Alert/Comments

The information from the SDRT may be useful for determining performance levels in decoding, vocabulary, comprehension, and reading rate. This information may be helpful for making instructional planning and evaluation of student progress decisions in reading. Technical information, which is needed for scoring and interpretation, is highly sophisticated and requires expertise in reading diagnosis and evaluation. The SDRT is well

standardized and reliable enough to ascertain specific areas of reading in which pupils display skill-development strength and weakness. When used as a criterion referenced test, certain subtests maybe useful for assessing LEP students and linguistically and culturally diverse students whose development of English language skills has been determined. Caution should be used in interpreting results because of possible cultural loading of reading passages. Scores may not be reported for bilingual and/or LEP students.

References

Manual
Salvia & Ysseldyke, 1985

Publisher

The Psychological Corp.

Title	<u>Test of Reading Comprehension (TORC)</u>	Age	<u>7-18 years</u>
Author	<u>V. L. Brown, D.D. Hammill, & J. Wiederholt</u>	Time	<u>Untimed (approx. 2 hours)</u>
Copyright Date	<u>1986</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The TORC, the successor to the original TORC published in 1978, is purported to assess silent reading comprehension in specific academic and academic-related areas.

Description

The TORC, which is individually and/or group administered, consists of eight subtests in a multiple-choice format. The General Comprehension Core consists of General Vocabulary, Syntactic Similarities, Paragraph Reading, and Sentence Sequencing. The four diagnostic supplements are Mathematics Vocabulary, Social Studies Vocabulary, and Science Vocabulary, and Reading Directions which is designed for younger and remedial readers.

Examiners should be familiar with administration and scoring procedures. . The manual includes clear and simple instructions and technical data. Ceiling levels are provided to shorten administration time.

The test kit includes the manual, Student Booklets, Student Answer Sheets, and Summary Profile Sheets.

Norm Data

The standardization sample consisted of 2,492 students, ages 7 to 18 years, from 13 states throughout the U.S. The sample was stratified according to sex, place of residence, and geographic distribution in accordance with the 1985 U.S. Census.

Results for individual subtests are reported as raw scores, percentiles, and standard scores. A Reading Comprehension Quotient (RCQ) is calculated from the General Comprehension Core score.

Reliability

Internal consistency for the TORC subtests was high; with one exception, coefficients were generally greater than .80; RCQ coefficients, ranged from .89 to .96. Test-retest reliability, based on retesting 59 students over a 1-week interval, yielded coefficients ranging from .64 to .86; RCQ was .91.

Validity

Content validity was developed as a result of item selection procedures. Construct validity was indicated by: positive correlation of scores with chronological age, the intercorrelation of subtests, the high correlation with tests of intelligence and/or academic achievement; and the capacity to differentiate between students who read well and those who have serious difficulty in reading. Criterion related validity was demonstrated by comparing performance on the TORC with the CAT, the Comprehensive Test of Basic Skills, the SRA Achievement Series, the PIAT, and the WISC-R; only two of the 32 coefficients reported were not significant at the .05 level.

Special Alerts/Comments

The TORC is designed to assess silent reading comprehension. The information obtained may be useful for determining performance levels, making instructional planning decisions, determining program effectiveness, and developing short- and longterm objectives on the IEP in the area of reading.

According to the test manual, "the test format is suitable for students with articulation problems and for those who are deaf, because spoken syntactic or morphological deviations from the text are not penalized" (p. 15).

Due to possible cultural loading of reading passages, caution should be used in interpreting results with linguistically and culturally diverse students. Its use with LEP students has not been demonstrated. Scores may not be reported for bilingual and/or LEP students.

Reference

Manual

Publisher

PRO-ED

Title	<u>Woodcock Reading Mastery Tests</u>	Grade	<u>K-12th</u>
Author	<u>R. W. Woodcock</u>	Time	<u>Untimed (20-30 minutes each subtest)</u>
Copyright Date	<u>1973</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The Woodcock is designed to assess skill development in reading.

Description

The Woodcock is a battery of five individually administered tests. The tests are contained in an easel format. The five subtests provide individual scores and a total reading performance score. Students are required to pronounce either names and words in isolation, use phonic and structural analysis, provide word meaning using analogies, and use a modified cloze procedure to indicate understanding of passage meaning.

It is suggested that a basal level, determined by five consecutive errors, be established to minimize the number of items which must be administered.

The examiner should calculate the raw score for each subtest before administering the next test. The student's total reading score is obtained by converting the raw score of each subtest to mastery scores and averaging the mastery scores. A table supplying conversion scores is presented in the Manual. No additional or special materials are required for administration.

The five skill areas of reading addressed by the Woodcock are: letter identification, word identification, word attack, word comprehension, and passage comprehension.

Norm Data

A total of 5,252 students was used in the standardization sample which was representative of the U.S. population in terms of race, years of schooling, occupation, income, and community size. Detailed descriptions of the norm sample and its limits were clearly provided in the manual. Based upon performance on the Woodcock, a grade score, age score, percentile rank, standard score, and mastery score can be generated.

Reliability

Split-half reliability coefficients were provided for forms A and B in grades 2 and 7. With the exception of the Letter Identification subtest, reliability coefficients ranged from .79 to .99. Reliabilities were generally lower for grades 7 than for grade 2, especially for Letter Identification.

Validity

The author reported information on content, construct, and predictive validity. The content validity was based on test development procedures. Construct validity was indicated through a multimethod-multitrait matrix in which scores for a small sample of second and seventh grade students were correlated; results generally supported the criteria of validity. However, these very same data supported the concept of reliability, equivalence of alternate forms. Similarly, predictive validity over a one-week interval, based upon predicted success levels on an alternate form of the test, indicated a high relationship between predicted and observed test score levels.

Special Alerts/Comments

The information provided by the Woodcock is useful for determining performance levels, instructional planning, and evaluation of student progress in reading as well as information useful for developing short- and longterm objectives on the IEP. However, due to low reliability coefficients, subtests of the Woodcock should not be used in isolation but only as components in determining a total score for reading. The author reports that the total score is the most reliable index of skill development in reading.

A unique feature of the Woodcock is the Mastery Scale which provides the range of a student's reading proficiency at different grade levels of difficulty. Scoring of the tests requires several numerical computations which can be quite time consuming for the examiner. Interpretation of scores on word and passage comprehension should address the type of limited behavior being assessed (i.e., analogies and passage completion). The Woodcock does not follow the sequence of the MTEs.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously. Scores should not be reported for bilingual and/or LEP students.

Reference

Manual

Publisher

American Guidance Service, Inc.

LANGUAGE ARTS

Language arts measures include tests of spelling and writing. Some measures of language arts achievement may be found in the section on achievement batteries.

Title	<u>Test of Written English (TWE)</u>	Grade	<u>2nd - 6th</u>
Author	<u>V.R. Anderson & S. K. Thompson</u>	Time	<u>10 - 20 minutes</u>
Copyright Date	<u>1979</u>	Type	<u>CRT</u>

Description

The TWE is intended to assess written English.

Intended Purpose of Test

The TWE consists of a 4-page recording form and an accompanying manual. The first page of the recording form provides space for factual information about the student and a profile of results; the remaining three pages are used by the student. The manual provides administrative, scoring, technical, and instructional information.

The test items are arranged into four subtests: Capitalization, Punctuations, Written Expression, and Paragraph Writing. The Capitalization subtest contains 9 items; the Punctuation section contains 13 items; the subtest for for Written Expression consists of 6 parts, with a total of 17 items; and the Paragraph Writing subtest directs the student to complete a paragraph for which the first two sentences are presented. Items in each subtest are sequenced in order of difficulty. The level of mastery depends on the student's grade level.

Test Administration

The TWE is administered individually, requiring approximately 10 to 20 minutes. It may be administered on a small group basis only if each student is able to read.

It is recommended that testing be discontinued on the Capitalization and Punctuation subtests after two consecutive errors.

An answer key is provided in the manual, which also provides information about the total number of points possible for each item. An overall performance level may be obtained by averaging the grade scores for the first three subtests. This performance level may then be compared with that attained on the Paragraph Writing subtest.

Technical Information

The underlying assumption of the TWE is that written expression is one of the areas where a severe discrepancy may appear between a student's academic functioning and his/her intellectual capabilities.

Curriculum guides, language books, and readings were examined to develop the order and grade levels for each skill area.

Effective Use/Comments

Although the TWE may be useful for obtaining descriptive information about the student in written expression, the test findings are insufficient for making performance level and individual instructional planning decisions in the area of writing. Additional assessments are necessary. Information provided may be useful in developing shortterm objectives on the IEP in the area of written English.

The manual provides strategies for teaching written expression, arranged according to grade level. However, the contents of grade levels in the TWE manual do not necessarily follow the sequence of the MTE's.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

Reference

Manual

Publisher

Academic Therapy

Title	<u>Test of Written Language (TOWL)</u>	Age	<u>8-1/2 yrs. - 14-1/2 yrs.</u>
Author	<u>D.D. Hammill & S.C. Larsen</u>	Time	<u>Untimed (approx. 40 mins.)</u>
Copyright Date	<u>1978</u>	Type	<u>NRT</u>

Intended Purpose of Test

The TOWL is designed to provide a measure of a student's written expression in a contrived and a spontaneous writing format.

Description

The TOWL is either group or individually administered and consists of the following subtests: Vocabulary, Thematic Maturity, Spelling, Word Usage, Style, Thought Units (supplemental), and Handwriting (supplemental). To measure Vocabulary, Thematic Maturity, Spelling, Word Usage, Style, Thought Units, and Handwriting, the student is asked to create a story (minimum of 50 words) about three pictures. For Word Usage, the student must complete sentences; for Style, the student must correct sentences for punctuation and capitalization; for Spelling, the student writes dictated words. The results produce raw scores, grade equivalent scores, scaled scores, and a Written Language Quotient.

Norm Data

The sample consisted of 1,700 unselected students, ages 8.5 through 14.5, living in nine states. Residence (urban/rural), sex, and geographic area were provided in the manual. No details regarding race and socio-economic level were provided.

Reliability

Three types of reliability were reported: internal consistency, test-retest, and interscorer reliability. Internal consistency coefficients were reported for Style, Word Usage, and Spelling only, and ranged in the .80's. Test-retest reliability, using a two-week interval with 50 students, ranged from .62 to .94. Interscorer reliability, indicated as indices of agreement, which were based upon scores obtained from 15 experienced teachers, was reported as .93 for Thematic Maturity, .97 for Thought Units, .76 for Handwriting, and .98 for Vocabulary.

Validity

Three types of validity were reported: content, construct, and concurrent. Content validity was based on test development and item analysis. Concurrent validity, using the Picture Story Language Test, yielded coefficients ranging from .16 to .60. Comparisons of scores with informal teacher ratings yielded correlations of greater than .50 except for Vocabulary and Handwriting. To provide data on construct validity, correlations of subtest scores and chronological age were computed; all r 's were significant and ranged from .18 to .55.

Special Alerts/Comments

The TOWL provides a measure of a student's written expression. The results of the TOWL may be useful for making screening decisions about written language. In conjunction with other language measures, it may be helpful for instructional planning. According to the authors, it is not appropriate to determine specific linguistic or semantic deficits, learning disabilities, language impairment, and/or dysgraphic.

While the test is easy to administer, the scoring and interpretation of the story is difficult and requires considerable training. The TOWL is a unique measure of written language and requires further systematic study to determine its full use in educational planning.

Given the task demands of the TOWL, it is not appropriate for LEP students and mentally retarded students.

References

Manual
McLoughlin & Lewis, 1981
Salvia & Ysseldyke, 1985

Publisher

PRO-ED

Title	<u>Test of Written Spelling (TWS)</u>	Grade	<u>1st - 8th</u>
Author	<u>V.R. Anderson & S. K. Thompson</u>	Time	<u>Untimed (10-20 minutes)</u>
Copyright Date	<u>1976</u>	Type	<u>NRT</u>

Intended Purpose of Test

The TWS is intended to assess spelling competency.

Description

The test consists of two categories of words: 35 "predictable" words, which are rules-governed and 25 "unpredictable" words which do not conform to any rules. The examiner administers the test by first saying the word in isolation, then in a sentence, and again in isolation. The student is required to write the words. The test yields three scores: Spelling Age (SA), Grade Equivalent (GE), and Spelling Quotient (SQ). The test can be administered to groups. However, individual administration is recommended.

Norm Data

The test was standardized on an unselected sample of 4,544 students from 22 states. The characteristics of the students relative to location, place of residence (cities-rural), social status (estimated by their teachers), grade level, and age were described in the manual: 78% were from the middle socioeconomic level as compared with 8% to 15% for the lower and upper levels, respectively. No information was provided about racial/ethnic background and class placement.

Reliability

Internal consistency coefficients (KR21) were computed for grade levels 1 through 8 and were all above .80 with the exception of grade 1 (.78). No other reliability data were presented.

Validity

Content validity with respect to "predictable" and "unpredictable" words was based on a computer search of English spelling rules. The content validity is impressive. Concurrent validity was measured by the relationship between the TWS and four other tests measuring spelling: Durrell .90, WRAT .84, CAT .80, and SRA .69.

Special Alerts/Comments

The TWS provides a measure of a student's ability to spell "predictable" and "unpredictable" words. The results of the TWS may be useful for making instructional planning, including development of short- and longterm IEP objectives, and evaluation of student progress decisions. The information is insufficient for determining the level of performance.

The test has wide application for a variety of populations (the manual discusses possible test modifications). Interpretation is limited to ability to spell "predictable" and "unpredictable" words. The test does not provide a format for analyzing the full range of spelling errors. Although the authors indicated a "Spelling Quotient" is obtainable, its meaning and educational usefulness is questionable.

The test is clearly not appropriate for use with linguistically diverse and LEP students.

Reference

Manual
Salvia & Ysseldyke, 1985

Publisher

PRO-ED

MATHEMATICS

Mathematics tests include assessment of a student's proficiency in arithmetic and higher mathematical processes. Tests include sections on computation, concept development, problem solving, and measurement. Others include algebra, geometry, statistics, and/or probability. Graphs are included in some tests.

Title	<u>Design for Math</u>	Grade	<u>K - 6th</u>
Author	<u>K. Kamp, J. Armenia, D. McDonald, & L. Von Kuster</u>	Time	<u>1 Hour</u>
Copyright Date	<u>1979</u>	Type	<u>CRT</u>

Intended Purpose of Test

The Design for Math is designed to measure a student's competency in mathematics skills through pre- and posttesting.

Description

The Design for Math is a criterion referenced test consisting of 176 mathematics skills, arranged in seven levels of math ability. It contains pre- and posttesting components. Pretesting components include Placement Tests and Diagnostic Progress Tests. Placement Tests are available in three ability ranges and in either hand or machine scored format. Diagnostic Progress Tests are available in spirit master form in the first four levels and in reusable test booklets in the remaining levels. There are two equivalent forms for each test level.

- | | |
|---------|--|
| Level A | size; weight; liquid and dry measures; triangular, circular, square shapes; classifying and matching 3-D shapes and patterns; sequencing events; recognizing coins; objects and sets of objects; equivalent and nonequivalent sets; matching numerals and sets; counting. |
| Level B | inside, outside, on; surfaces, curves; name and draw shapes; sums/differences; cardinal numbers (0-9); sentences; length, liquid and dry measures; weight; simple graphs; day of the week, nearest hour; ordinals (1st-10th); numbers and sets (0-12); numbers (0-99); grouping by tens (0-100); number line (0-20); sums and differences (10-10); addition of three 1-digits; coin value, cent notation; equivalent sets; total of objects grouped; repeated addends (sums to 10), fractional models ($\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$). |
| Level C | fractions ($\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$); number line fractions; time to $\frac{1}{2}$ hour; grouping objects; numerals for grouping objects; numerals for grouped sets of tens; missing addends; addition and subtraction facts; word problems; sums/differences (100 or less); regrouping; commutative property of addition; counting (2s, 3s, 4s, 5s, 10s); equal addends-multiplication equations; addition/subtraction with coins; change to a dollar; ordering numbers (to 100), days of the week; elapsed time in days; weight, liquid and dry measures; temperature (Celsius); identifying line segments; length (centimeter, inch); picto, bar, circle graphs. |

- Levels D symmetry; congruence; reflections; radius and diameter; perimeter; weight (kilograms, pounds); temperature (Fahrenheit); associative property of addition; subtraction of 3-digit numbers, transactions and computations (to \$10.00); commutative property of multiplication; multiplication facts; missing factors; word problems, fractions (through eighths); broken line graphs; mixed numbers-number line; time (to nearest five minutes); area (nonstandard units); liquid measures; solid volume (nonstandard units); odd and even; compact form; equality and inequality signs.
- Level E time (to nearest minute); lines and angles; properties of circles; weight (in ounces and grams); liquid and dry measures; rounding whole numbers (to 100,000); constructing and interpreting graphs; prime and composite numbers; ordering fractions; addition/subtraction of fractions (with like denominators); length (metric and customary); elapsed time (minutes-years); products (to 81); missing factors; divide 2-digits by 1-digit, ($r=0$); word problems; area; computation of volume; differences of 4-digit numbers; estimating sums (4-digit numbers); compute transactions; multiply 3-digits by 1-digit; divide 2-digits by 1-digit; associative property of multiplication.
- Level F representation of solids; measure and attributes of angles; length (metric, customary); compute areas; surface areas and volume; multiply 3-digits by 2-digits; divide 4-digits by 1-digit; prime factors; least common multiple; fractions (to twelfths); equivalent fractions; renaming (simplest form); mixed numerals; multiplication of fractions; unlike denominators; division of fractions; whole numbers to (1,000,000); integers; bases other than ten; prime and composite numbers (1-100); appropriate measurement instruments; metric multiples and submultiples; sums and differences; word problems; elapsed time (seconds-years); graphing (positive rational numbers); points on a coordinate grid.
- Level G whole numbers (to 1,000,000,000); Roman numerals; sequence patterns; bases other than ten; multiply 5-digits by 3-digits; area of a circle; measurement problems; measurement instruments and units; prisms and cylinders; metric mass and volume; division of 5-digits by 3-digits; compute average; constructing plane figures; lines and planes in space; graphic linear equations; interpreting graphs found in media; computations with fractions and mixed numbers; division with mixed numbers; decimals (through hundred thousandths); addition/subtraction, multiplication/division of decimals; powers to five; addition/subtraction of integers; word problems; consumer money matters; time zones; time; rate and distance problems; distance from scale drawings.

Test Administration

Design for Math may be group or individually administered. Most subtests are formal paper and pencil tests. The manual indicates, however, that some math skills cannot be assessed effectively by formal testing and specific directions are, therefore, included for assessment by Teacher Observation performance tests. Mastery ratios for each subtest are provided in the manual. No special training or materials are necessary.

Technical Information

Design for Math was developed in response to educators' expressed need for a manageable and practical means of organizing existing math curriculum. It is modeled after the Wisconsin Design for Reading Skill Development. The manual states that there are no norms for pre- and posttests. No information is provided pertaining to the method used to determine mastery ratios.

Effective Use/Comments

Design for Math is designed to measure a student's competency in mathematics skills. Given the lack of technical information, results are insufficient for making screening, performance level, instructional planning, and evaluation of student progress decisions. With other mathematics tests, the information provided is sufficient for developing long- and shortterm objectives on the IEP in math and for evaluating student progress.

Design for Math may be used with LEP students or linguistically and culturally diverse students with caution.

Reference

Manual

Publisher

NCS/Educational Scoring Systems

Title	<u>Diagnostic Mathematics Inventory (DMI)</u>	Grade	<u>1.5 - 8.5</u>
Author	<u>J. Gessell</u>	Time	<u>Untimed</u>
Copyright Date	<u>1978</u>	Type	<u>CRT</u>

Intended Purpose of Test

The DMI is designed to measure a student's mastery of mathematics skills normally taught in grades 1-8. It is useful for diagnosis or instructional planning and evaluation of instruction.

Description

The DMI is a criterion referenced test consisting of a series of multiple-choice tests assessing mathematical skills, divided into seven levels:

A/Red 1.5 to 2.5	Counting; matching; addition and subtraction of whole numbers measurement; sequences; missing addends; geometry; numeration systems.
B/Green 2.5 to 3.5	Counting; regrouping in addition and subtraction; multiplication; expanded measurement; preoperational concepts in fractions.
C/Blue 3.5 to 4.5	All basic operations with whole numbers and their properties; expanded multiplication; division; fractions; geometry; measurement; place value; problem-solving.
D/Orange 4.5 to 5.5	All basic operations with whole numbers and their properties; fractions, decimals; measurement; geometry (lines, rays, angles, Cartesian coordinates).
E/Aqua 5.5 to 6.5	All basic operations with whole numbers and their properties; fractions; decimals; number theory, measurement; place value; geometry; problem solving.
F/Purple 6.5 to 7.5	All basic operations with whole numbers and their properties; fractions; decimals; number theory; measurement; place value; geometry; problem solving; computation of negative integers; sets; percent; statistics.

G/Brown 7.5 to 8.5

All basic operations with whole numbers and their properties; fractions; decimals; number theory; measurement; place value; geometry (points, lines, triangles, parallels, perpendiculars, polyhedra, ellipses, Cartesian coordinates); problem solving (mathematical sentences); computation and properties of negative integers; operations with sets; percent; statistics; numeration systems.

Learning activities guides which match specific test items and objectives provide instructional activities for teaching these objectives. A Guide to Ancillary Materials correlates objectives with a variety of nontest materials; Master Reference Guides correlate objectives to text materials, referencing specific textbook pages to each objective.

Test Administration

The DMI may be administered individually or in groups. No special training or materials are required. It is recommended that each test session be limited to no more than one hour at a time.

Technical Information

The manual for the DMI contains no technical information. However, comparison of test specifications with MTEs indicates the test has content validity for New York City. The test measures what should be taught in grades 1 through 8. Each test item represents the objective with which it is identified.

Although criteria for mastery are given, the method for obtaining these criteria was omitted.

Effective Use/Comments

The information provided by the DMI is sufficient for developing short- and longterm objectives on the IEP. Although technical data are lacking, Salvia and Ysseldyke (1985) noted the test's completeness and recommended it for instructional planning. - Nonetheless, the information provided is not sufficient for making screening, and performance level decisions in the area of mathematics. Materials include specific strategies regarding instruction.

The DMI may be used with linguistically and culturally diverse students and LEP students.

References

Manual
Salvia & Ysseldyke, 1985

Publisher

CTB/McGraw Hill

Title	<u>Enright Diagnostic Inventory of Basic Arithmetic Skills (EDI)</u>	Grade	<u>1st - 12th</u>
Author	<u>B. E. Enright</u>	Time	<u>Untimed</u>
Copyright Date	<u>1983</u>	Type	<u>CRT</u>

Intended Purpose of Test

The EDI is designed to measure competency in basic arithmetic computation skills for students and to diagnose student arithmetic needs as a result of error analysis. It also provides an instructional management system.

Description

The EDI is published in a sturdy looseleaf binder with a convenient flip-page arrangement for one-to-one administration. It consists of group administered Placement Tests, Skill Tests, and Basic Facts Tests on the addition, subtraction, multiplication, and division of whole numbers, fractions, and decimals.

The Wide Range Placement Test, available in two forms, screens for the operations of addition, subtraction, multiplication, and division of whole numbers, fractions, and decimals. As a result of information obtained from the Wide Range Placement Test, Skill Placement Tests are given to determine the extent of a student's knowledge in an arithmetic operation and to identify the need for more precise skill testing. There are thirteen Skill Placement Tests, available in two forms; 100% mastery is required on each test. The Basic Facts Tests identify students whose fundamental weakness lies in computing the basic facts. Each of the four basic arithmetic tests, available in two forms, consists of 50 items; 98% accuracy is required.

As a result of information obtained from the above tests, an appropriate skill test is selected for the diagnosis of student needs. A remedial instructional program is developed for each student based upon an analysis of errors on the skill test. A total of 144 subskills has been identified in the addition, subtraction, multiplication, and division of whole numbers, fractions, and decimals. A skill test has been developed for each subskill; these are sequenced from the simplest to the most complex. Each test page consists of a pretest and posttest of five items; 90% accuracy is required.

Additional components include an Arithmetic Book and a Class Record Sheet which offer a simple system for recording student and group performance and for identifying individual error patterns. An optional record keeping system of forms, the Individual Progress Record, in duplicate, is available for whole numbers, decimals, and fractions; these may be used for IEP development and parent conferences.

Test Administration

Although the EDI was designed for one-to-one administration, many of the subtests can be group administered by an examiner who needs no special training.

The examiner gives the Skill Placement Test, determined by the Wide Range Placement Test or by the student's current classroom performance, prior to administering Skill Tests to obtain diagnostic information. Diagnosis is accomplished by using the Error Analysis which is included on the examiner's page. Error Analysis enables the assessor to determine the specific error pattern made by the student, such as regrouping, process substitution, omission, direction, placement, attention to sign, and guessing. Remediation activities are designed through error analysis.

Posttests should be administered before proceeding to more complex skills.

Technical Information

The instrument was developed and field tested over a period of five years, including a pilot by 100 teachers with thousands of students in New Orleans, Louisiana for "several months". Content validity, based upon feedback from 40 educational specialists and university faculty in educational assessment, was developed during field test revision; items for which there were 90% agreement were retained in the final copy of the test.

Based upon small samples of students, split-half reliability coefficients ranging from .91 to .98, were reported for each of the thirteen Skill Tests.

No information was presented on how mastery was determined.

Effecton Use/Comments

The EDI provides information on arithmetic computation skills which is not sufficient for making performance level decisions. The information provided by the EDI is sufficient for instructional planning and developing short- and longterm objectives on the IEP in arithmetic computation.

The inventory was not intended to be inclusive; word problems, money, and measurement were omitted. The tests closely follow the sequence of the MTEs. The instructional objectives are specific. The error analysis provides the teacher with precise diagnostic information. However, the manual does not include specific strategies for instruction.

The instrument may be used with linguistically and culturally diverse students and LEP students.

Reference

Manual

Publisher

Curriculum Associates

Title	<u>Key Math Diagnostic Arithmetic Test</u>	Grade	<u>K - 8th</u>
Author	<u>A. Connally, W. Nachtman, & E. Pritchett</u>	Time	<u>Untimed (30 minutes to 1 hour)</u>
Copyright Date	<u>1971, 1976</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

Key Math is designed as both a norm referenced and criterion referenced measure to assess arithmetic skill development in kindergarten through eighth grade.

Description

Key Math is contained in an easel kit which presents stimulus materials to the student at eye-level, while the examiner's instructions are placed on the reverse side. It is comprised of 209 items, organized into 14 subtopics related to content, operations, and applications, and is individually administered. Many of the subtests require an oral response from the student. Mastery is met by obtaining the correct response.

The entire test is not given to each student; a basal level, established by three consecutive correct responses, and a ceiling level, marked by three consecutive errors, are identified.

<u>Area</u>	<u>Subtest</u>	<u>Skills</u>
Content	Numeration	identifying numerals; one-to-one correspondence; counting; comparing sets; numerical sequencing; ordinal numbers; Roman numbers; recognizing decimals; squared numbers; ratio
	Fractions	pictorial concept of whole and half; numerical concept of half; relating parts to the whole; reading fraction symbols; mixed numbers
	Geometry & Symbols	identifying and discriminating among geometric shapes; combining forms into common geometric shapes; defining operational symbols; mathematical abbreviations; parallel and perpendicular lines
Operations	Addition	visual addition problems presented and answered orally; written computation addition problems
	Subtraction	oral subtraction problems, written subtraction problems including decimals, money, and mixed numbers

	Multiplication	problems presented and answered orally; written problems involving money.
	Division	problems presented and answered orally; written problems, one involving money and one involving a fraction to be reduced to lowest terms.
	Mental Computation	simple to increasingly difficult single and multi-operational problems to be done without pencil and paper.
	Numerical Reasoning	horizontal operational problems where a symbol represents a number missing from the problem or its solution.
Applications	Word Problems	problems presented and answered orally with visual representations, presented orally with only important information printed on subject's side of easel; presented orally with whole problem printed on student's side of easel.
	Missing Elements	orally presented problems containing one missing piece of information, some with visual representations, some with only important information on student's side of easel, some with whole problem printed on student's side of easel.
	Money	oral problems with visual representations including coins, budget graph, check, dollar bills, check register, bank statement.
	Measurement	ruler; outdoor thermometer; balance scale, and oven thermostat; meanings of liter, volume and weight measurements; eye-measurement; perimeter and area.
	Time	clock, season, calendar, year and decade, with pictorial representations.

Norm Data

Key Math was standardized in 1971 on a group of more than 1,222 students in kindergarten to seventh grade. Students were from 42 school districts in eight states. The sample contained students from several geographic areas, several races, and from rural, urban, and suburban communities. No information was presented in the manual about sex or handicapping conditions. The normative data were weighted to conform to the proportions obtained on race and community size as described in the 1970 U.S. Census.

Reliability

Reliability coefficients, reported by subtest and grade level (K-7), ranged from .36 to .90, with the median reliability across subtests and grades ranging from .64 to .84. Test-retest reliability based on a Minneapolis school study with EMR students, yielded coefficients ranging from .24 to .74 for a 6-month interval.

Validity

Comparing Key Math performance of 28 "normal" fifth graders to their performance on the Iowa Test of Basic Skills (ITBS), yielded a concurrent validity correlation of .38. The authors admitted that inadequate data exist for demonstrating concurrent validity.

Special Alerts/Comments

Key Math provides information about a student's abilities in arithmetic. Given its limited technical development (Salvia & Ysseldyke, 1985, p.379), the information provided is insufficient for making performance level and evaluation of student progress decisions.

When used as a criterion referenced test, the information provided by the Key Math is sufficient for making planning decisions in arithmetic and for developing short- and longterm objectives on the IEP in this area. Since the objectives are linked to one test item only, further assessment is required to determine student needs.

Since some subtests have high verbal loadings, its use with linguistically and culturally diverse students and LEP students should be done cautiously. Scores may not be reported for bilingual and/or LEP students.

References

Manual
McLaughlin & Lewis, 1981
Salvia & Ysseldyke, 1985

Publisher

American Guidance Service, Inc.

Title	<u>Stanford Diagnostic Math Test (SDMT)</u>	Grade	<u>1.5 - 12th</u>
Author	<u>L. Beatty, R. Mudden, E. Gardner, & B. Karlisen</u>	Time	<u>1-1/2 - 2 Hours</u>
Copyright Date	<u>1976</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The SDMT is designed to measure competence in basic mathematical concepts and skills and identify a student's specific strengths and weaknesses in mathematics. It is specifically intended for students in regular and special education programs who are performing below average in mathematics.

Description

The SDMT is both a group administered norm referenced and criterion referenced test. The SDMT provides four color coded levels of diagnostic evaluation with two parallel forms (A&B) at each level. All four levels assess three interrelated areas of mathematical competence described as the Number System and Numeration (Concepts), Computation, and Application. At each level, the manual provided instructional objectives.

Test 1 <u>Number System and Numeration</u>	whole numbers and decimal place value; rational numbers and numeration; operations and properties
Test 2 <u>Computation</u>	addition, subtraction, multiplication and division; fractions; decimals; percent; number sentences
Test 3 <u>Application</u>	problem solving; reading and interpreting tables and graphs; geometry and measurement

Criteria for mastery, called Progress Indicators, range from 33% to 75%.

Norm Data

The SDMT was standardized in 1975 on a stratified, random sample of 35,000 students in grades 2 through 12 resembling the national school population in terms of school system enrollment, ethnic groups, geographic area, urbanization, and socioeconomic status. There was no information regarding the breakdown of the sample as to sex and handicapping conditions. In addition to percentiles, scale scores, stanines, and grade equivalents for subtests and total, Progress Indicators (PI) have been developed for each concept/skill domain to indicate mastery of each instructional objective.

Reliability

Internal consistency coefficients of the SDMT ranged from .84 to .97 for the four levels of the test; alternate form coefficients ranged from .64 to .94 with most above .80. Tetrachoric correlation coefficients to determine alternate form reliability of the PI's ranged from .29 to .99, with most in the upper range.

Validity

Content validity was based on correspondence to the mathematics curricular content. Criterion validity was established by correlating performance of the SDMT with the mathematics subtest of the SAT. Coefficients ranged from .64 to .94.

Special Alerts/Comments

The SDMT provides information about a student's performance level in basic mathematical concepts and skills. The results may also be useful for making evaluation of student progress and instructional planning decisions including the developing of short- and longterm objectives on the IEP in the area of mathematical concepts and skills.

The test follows the sequence of the MTEs. In addition to specifying objectives, the manual provides specific strategies regarding instruction and instructional management.

The test may be given individually for handicapped students. Such students may record responses in the test booklet rather than on a separate answer sheet. This prevents errors in recording and allows the examiner to analyze the student's calculation procedures.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously since some of the subtests may have high verbal loadings. Scores may not be reported for bilingual and/or LEP students.

Reference

Manual
Buros, 1978
Salvia & Ysseldyke, 1985

Publisher

The Psychological Corp.

Title	<u>Test of Computational Processes</u> <u>(TCP)</u>	Grade	<u>1st - 8th</u>
Author	<u>N.D. Kingston</u>	Time	<u>Untimed (approximately 1 hour)</u>
Copyright Date	<u>1985</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The TCP is designed to measure competency in basic computational skills for regular, remedial, and special education students and to diagnose arithmetic needs as a result of error analysis.

Description

The TCP may be administered individually or to a group of students by a professional who is familiar with test procedures. The 8-page test booklet contains 122 items which measure computational skills in seven domains: adding, subtracting, multiplying, and dividing whole numbers; fraction and decimal computation; measurement (US and metric).

Three types of scoring procedures are available: General Scoring Standards determine right/wrong answers by assigning full, part, or no credit to test items. Process Scoring Standards are optional and only consider the use of proper algorithms (processes); items with careless errors may receive full credit if the correct algorithm was applied. Diagnostic Scoring provides an in-depth analysis and indicates content deficiencies and error patterns.

The examiner's manual includes administration/scoring procedures.

Norm Data

Standardization was based on more than 6,000 students, grades one through eight, from six school districts in Idaho and Wyoming. Testing was conducted in the Spring and Fall of 1982 and included an equal number of males and females; minority students comprised about 7% of the population.

Results are reported as raw scores, converted scores, percentile ranks and grade ratings for the total test (total score), nonmeasurement items (critical score), and computational items (process score).

Reliability

Split-half reliability, corrected by the Spearman-Brown formula, was extremely high; coefficients ranged from .87 to .99.

Validity

Content **validity** was established as a result of 12 years of test construction and **development**. Construct and criterion-related validity were not addressed.

Special Alerts/Comments

The TCP is designed to measure competency in basic computational skills. Given the limited sample, the information provided is insufficient for making performance level decisions. The information may be useful for making instructional planning decisions and developing long- and shortterm objectives on the IEP in the area of mathematics. This test closely follows the sequence of the MTEs.

The TCP may be used with linguistically and culturally diverse students and LEP students. Scores may not be reported for bilingual and/or LEP students.

Reference

Manual

Publisher

DLM Teaching Resources

Title	<u>Test of Mathematical Abilities(TCMA)</u>	Age	<u>8-1/2 - 19 years</u>
Author	<u>V. L. Brown & E. McEntire</u>	Time	<u>Untimed (45 minutes)</u>
Copyright Date	<u>1984</u>	Type	<u>NRT</u>

Intended Purpose of Test

The TOMA is designed to assess a student's strengths and weaknesses in mathematic and mathematically related abilities.

Description

The TOMA can be either group or individually administered. The test assesses mathematical abilities in story problems and computation, attitudes toward mathematics, understanding of vocabulary, and understanding the functional use of mathematics facts and concepts. The TOMA Math Quotient (TMQ) is used as a general index of the student's overall mathematical abilities. Basal and ceiling levels are established for all subtests except the Attitude toward Math subtest which must be given in its entirety. It is recommended that the Attitude toward Math subtest be given first so as to provide a student's general attitude toward math. The TOMA can be administered in more than one session.

Norm Data

The TOMA was standardized on a sample of more than 1,500 students in diverse geographic areas of the U.S., stratified by sex, race, and type of residence (urban-rural), based on the 1980 census.

Raw scores, standard scores for all the subtests, percentiles, and a quotient (TMQ) that incorporates standard scores for all the subtests can be ascertained.

Reliability

Mean correlations using the coefficient-alpha procedure were at or above .80, which supported the internal consistency of the TOMA. Test-retest reliability coefficients, using a 6-week interval, ranged from .71 to .87 for various subtests. Test-retest reliability for the total test was found to be .80.

Standard errors of measurement are reported, thereby establishing ranges within which the true score probably lies.

Validity

Criterion-related validity was established by correlating the TOMA with the Key Math Test, the PIAT and the WRAT. Construct validity was demonstrated by investigating the

TCMA in term of age and grade differentiation, IQ, intercorrelation of subtests, and differentiation of students into groups known to have difficulties in math and groups thought to be normal.

Special Alerts/Comments

The TCMA provides information about a student's ability in mathematics which may be useful for making screening, performance level, and evaluation of student progress decisions. Because of the high verbal loadings on some subtests, the use of the TCMA is not recommended for poor readers, deaf, and LEP students. Students who read well may read Attitude toward Math to themselves.

The use of the TCMA with linguistically and culturally diverse students and LEP has not been demonstrated. Scores may not be reported for bilingual and/or LEP students.

Reference

Manual

Publisher

PRO-ED

ADAPTIVE BEHAVIOR

Adaptive behavior refers to how well individuals adapt themselves to the expectations of their environment. Adaptive behavior generally includes the areas of social competence and non-academic developmental tasks.

Adaptive behavior measures are not self administered nor are based upon direct performance; they generally result from observations or structured interviews which are conducted with parent(s), teacher(s), or others who are familiar with the student. The latter consist of third-party ratings of the student's ability to deal with a variety of social and familiar experiences. More direct measures of the student's functioning are necessary to obtain a total picture of the student's adaptive functioning; no decisions should be made in this area without direct measurement of student performance.

Adaptive behavior scales may be norm referenced or criterion referenced. Norm referenced scales are usually screening devices, comparing the student with his/her peers. Criterion referenced scales are usually indepth analyses of domains, such as daily living skills, socialization, or communication and language, which report results useful for developing short- and longterm objectives on the IEP.

Since the focus of adaptive behavior is on the performance of non-academic tasks, the scales may be appropriate for students from linguistically diverse backgrounds provided that language provisions are made by the interviewer. Cultural background must be considered in interpreting results since scores frequently show cultural differences rather than indication of a handicapping condition.

Title AAMD Adaptive Behavior Scale, School
Edition (ABS-SE) (Revised)

Age 3 - 16 years

Author K. Nihira, R. Foster,
M. Shellhaas, & H. Leland

Time 1 hour

Copyright Date 1981

Type NRT, CRT

Intended Purpose of Test

The ABS-SE is intended to assess an individual's adaptive behavior which is defined by the AAMD as "the effectiveness or degree with which a person meets the standards of personal independence and social responsibility expected for his or her age and cultural group" (Manual).

Description

The scale is a third-party behavior rating measure consisting of 95 items presented in two parts. Part One is designed to evaluate a person's skills and habits in nine behavior domains important to the development of personal independence in daily living: independent functioning, physical development, economic activity, language development, numbers and time, prevocational activity, self-direction, responsibility, and socialization.

Part Two measures adaptive behavior related to personality and behavior disorders: aggressiveness, antisocial vs. social behavior, rebelliousness, trustworthiness, withdrawal vs. involvement, mannerisms, interpersonal manners, vocal habits, activity level, symptomatic behavior, and use of medications.

The test is generally administered to teachers but the authors recommended that parents also be encouraged to fill out the ABS-SE for a complete record. Percentile distributions for each age and placement group were developed.

Norm Data

The sample for the 1975 version consisted of 2,600 California school children, age 7 to 13 years, classified as regular, EMR, or TMR students. In addition, the 1981 sample consisted of 3,220 regular, EMR, and TMR students, age 3 to 16 years, from Florida and 363 pre-school children, age 3 to 6 years, from six school districts in urban, suburban, and rural areas in California and Florida. Additional data were obtained from 656 California students in EMR and TMR classes who were re-evaluated and kept in special education classes. The samples were controlled for SES, sex, and ethnicity. Separate norms were developed for male/female and ethnic groups for Part Two. Norms are also available for Spanish students.

Reliability

Estimates of internal consistency obtained by the coefficient alpha technique ranged from .34 to .95 for the 21 domains, with a mean

reliability of .86. Among EMR and TMR populations, coefficients were generally highest for personal-social responsibility, social adjustment, and community self-sufficiency (.79 to .95). The low reliabilities of the domain scores tend to reduce the usefulness of the measure for making individual decisions (Salvia & Ysseldyke, 1985).

Validity

Examination of adaptive behavior and intelligence indicated weak to moderate relationships; in general, Part One domains correlated significantly with IQ but not with achievement tests, whereas Part Two domains were independent of both IQ and achievement tests. A study of predictive validity resulted in significant but low correlations between scores and classification (regular or EMR).

Special Alerts/Comments

The ABS-SE provides a gross measure of an individual's adaptive behavior which may be useful in obtaining descriptive information about a student's performance. However, as with other third-party rating scales, the ABS-SE is not a direct measure of student performance and, consequently, judgment of adaptive behavior requires additional direct behavioral information. In conjunction with other measures, the ABS-SE may be useful for planning instruction and developing short- and longterm objectives on the IEP in the area of adaptive behavior.

Doucette and Freedman (1980) questioned the use of Part Two in public school settings due to "lower reliability, a negative value-biased tone, item content that includes infrequently occurring behaviors, and skewed distributions" (p. 66). There should be cautious interpretation of domain scores based on a single set of norms. Sex, ethnic status, and linguistic background should be considered when interpreting scores. Elliott who reviewed the instrument for the most recent Mental Measurements Yearbook, reported that the ABS-SE is far superior to any other test of adaptive behavior currently on the market" (Mitchell, Jr., 1985 p.3).

Since the ABS-SE focuses on adaptive behavior, it may be appropriate for students from culturally and linguistically diverse backgrounds and LEP students provided that language provisions are made for the interviewer.

References

Manual
Doucette & Freedman, 1980
Lambert, Windmiller,
Cole & Figueroa, 1975
Mitchell, Jr., 1985
Salvia & Ysseldyke, 1985

Publisher

CTB/McGraw Hill

Title	<u>Adaptive Behavior Inventory for Children (ABIC)</u>	Age	<u>5 - 11 years</u>
Author	<u>R. Mercer & J.F. Lewis</u>	Time	<u>3 hours</u>
Copyright Date	<u>1977</u>	Type	<u>NRT</u>

Intended Purpose of Test

The ABIC is designed to secure information about the student's adaptation to the social systems in which he or she participates. The ABIC measures the student's social role performance in the family, the peer group, and the community. It was designed to be part of The System of Multicultural Pluralistic Assessment (SOMPA).

Description

The ABIC consists of 242 interview questions organized into six scales: Family, Community, Peer Relations, Non-Academic School Roles, Earner/Consumer, and Self-Maintenance. Administration consists of an interview with the principal guardian of the student. The interview can be conducted in English or Spanish.

Norm Data

The ABIC was standardized on a sample of 2,085 California public school students 5-0 years to 11-11 years of age. The students were randomly chosen from pre-selected school districts. Black, Hispanic, and white students were represented in approximately equal numbers.

Reliability

Split-half (odd-even) reliabilities for the scaled scores were calculated for each age level, and separately for each ethnic group. Reliabilities for the total score for the complete sample and for each ethnic group were .95 or above. For the separate scales, reliabilities ranged from .78 to .92 with a median of .86. Recent studies (Oakland & Feigenbaum, 1982) reported high internal consistency (.90 to .96) for each of the six scales. High interrater reliabilities have also been reported (Salvia & Ysseldyke, 1985).

Validity

Recent validity studies (Oakland & Feigenbaum, 1982) indicated that the ABIC is measuring adaptive behavior across different ethnic groups and is not greatly affected by IQ and achievement. Correlations between the ABIC and the WISC-R for white and black students in the standardization sample yielded coefficients of .16 with the Verbal Scale, .14 with the Performance Scale, and .17 with the Full Scale IQ.

Special Alerts/Comments

The ABIC provides a measure of the student's social role performance in the family, the peer group, and the community. However, as with other third

party scales, the ABIC is not a direct measure of student performance and consequently judgment of social competence will require direct behavioral information. With other measures, the findings may be useful for making instructional planning decisions and developing short- and longterm objectives on the IEP in the area of adaptive behavior. In order to interpret ABIC results correctly, training is advised.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

References

Manual
Oakland & Feigenbaum, 1982
Salvia & Ysseldyke, 1985

Publisher

The Psychological Corp.

Title	<u>Balthazar Scales of Adaptive Behavior</u> <u>(BSAB)</u>	Age	<u>All Ages</u>
Author	<u>E.E. Balthazar</u>	Time	<u>Untimed</u>
Copyright Date	<u>1973 (BSAB-II), 1976 (BSAB-I)</u>	Type	<u>CRT</u>

Intended Purpose of Test

The BSAB is designed to measure the self-help skills and coping behavior of profoundly and severely retarded students.

Description

The BSAB consists of two separate sets of scales: (1) the Scales of Functional Independence (BSAB-I) are used to measure self-help skills such as eating, dressing, toileting; (2) the Scales of Social Adaptation (BSAB-II) are designed to assess a wide variety of social functions. Each scale consists of subscales divided into individual behaviors which are rated through observation in the natural environment.

BSAB-I: Scales of Functional Independence

<u>Scale</u>	<u>Description</u>
Eating	Measures specific eating skills which are divided into five groups
Dressing	Measures levels of dressing skills which are elicited by the rater (six male and eight female activities)
Toileting	Measures toileting behaviors

BSAB-II: Scales of Social Adaptation

<u>Scale</u>	<u>Description</u>
Unadaptive Self-directed Behaviors	Evaluates inappropriate behaviors which represent social isolation
Unadaptive Interpersonal Behaviors	Includes inappropriate responses to interpersonal contacts such as aggression and withdrawal

Adaptive Self-directed Behaviors	reflects reasonable degrees of attentiveness, exploration, regard of one's body, and recreational activities
Adaptive Interpersonal Behaviors	Includes pre-communication behaviors, as well as non-verbal social vocalization and gestures
Verbal Communication	Measures inarticulate and articulate verbalizations of single words and expressions
Play Activities	Evaluates observed object-related activities as well as playful contact with others
Response to Instruction	Behaviors range from simple group or individual instructions to complex directions involving willingness to help others
Checklist	Includes personal care and other observable behaviors

Test Administration

The rater must be well versed in the behavior of the target population. He must be able to rate the degree of the assessed behavior on a scale from one to ten (from "never" to "all the time"). On the Eating Scale, the rater must be able to observe behavior closely without interfering in any way with the behavior itself. On the Dressing Scale, the subject must be able to respond to simple verbal or gestural instructions. The Toileting Scale is derived from interviews with an aide familiar with the subject. For the Social Adaptation Scales, trained raters must be able to make frequency counts of the target behaviors per unit of time.

Technical Information

The key to reliability of the BSAB is rater training. The manual specifies five common errors among raters who are not adequately trained. Both inter- and intrarater reliability must be considered. It was suggested that videotapes be utilized to judge intrarater reliability over time. In regard to validity, it is noted that the behavioral observations constitute the criterion measure itself. Therefore, it is stated that it is not relevant to inquire if the observer's judgments are valid. The relevant question is whether the observer has had sufficient training and experience to rate the behaviors of the retarded population. Data are given both for intrarater reliability over time and for the agreement of the rating of two observers working simultaneously.

Effective Use/Comments

The BSAB is designed to measure self-help skills and coping behavior of profoundly and severely retarded students. The information provided is sufficient for making evaluation of student progress and instructional planning decisions and developing short- and longterm objectives on the IEP in the area of adaptive behavior. With other information, results are useful for making performance level decisions.

The BSAB provides a very comprehensive system for assessment and program development. The immediate application of the scales to program planning and evaluation is a strong feature.

Rater training of an extensive nature is essential. For school administration, a natural setting within which the behaviors freely occur needs to be available.

Given these factors, the BSAB could be a useful and detailed programming tool. Comprehensive training instructions are included in the manuals. Use of this instrument for assessment alone would not seem to warrant the time needed to be invested. However, if program development is the goal, then the use of the BSAB would seem to be worth the time of training and administration.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

Reference

Manuals

Publisher

Consulting Psychologists Press, Inc.

Title Behavior Rating Instrument for Autistic and Other Atypical Children (BRIAAC) Age/Grade All

Authors B.A. Rutterberg, B.J. Kalish, C. Wevar, & E.G. Wolf Time 2 Hours or more

Copyright Date 1977 Type CRT

Intended Purpose of Test

The BRIAAC is designed to assess the present levels of functioning of very low functioning atypical and autistic children in eight key areas: relationship, communication, drive for mastery, vocalization and expressive speech, sound and speech reception, social responsiveness, body movement, and psychobiological development.

Description

The BRIAAC consists of eight scales which are divided into levels. The highest level is comparable to that of a normal 3-1/2 to 4-1/2 year old. A level of functioning based on a 10-point rating scale is indicated for each level within a scale through the use of unobtrusive observation.

Relationship to adult	Levels may vary from lack of response to empathic behavior.
Communication	The rating consists of how the child communicates, what the child communicates, and the differentiation of the child from others. Levels vary from lack of response to communication of what has occurred and/or what will occur.
Drive for mastery	Emphasis is on motivation or interest. Levels range from repetition to self-monitoring.
Vocalization and expressive speech	The observation is based on what is heard in terms of form and structure of speech, rather than content.
Sound and speech reception	The observation is based upon the child's awareness and reactions to sound in the environment. Levels range from lack of awareness to noise, voices, and music to more complex listening skills including following a 2-step direction.

Social responsiveness

Rating is not limited to the ability to carry out self-help functions such as toileting, eating, and dressing, but also includes interest in attaining such skills.

Body movement

Child is observed on one of two scales, active or passive. Choice of scale depends upon pre-assessment findings about child's type of movement behavior.

Psychobiological development

Based upon psychoanalytic theory, this scale purports to measure interest in bodily functions and sensations as they "evolve through biological maturation" (Manual, p. 39).

Various reporting forms are available depending on the purpose for assessment. The Individual Scale Score Sheet provides space for recording student scores and writing explanatory comments. The Total BRIAAC Score Sheet provides information for all scales on one form. The Intrascale Profile is a graphic presentation of each scale. The Interscale Profile may be useful for longitudinal comparisons of the scales; it is not appropriate for comparing scales because of the lack of comparability of scores. The BRIAAC Report Form contains a detailed description of the levels for each scale and provides space for the child's achievement.

Test Administration

Assessment is obtained through unobtrusive observation in the child's usual environment. The observer should be trained in the administration of the BRIAAC. It is recommended that the total BRIAAC Score Sheet be used for recording information. Detailed instructions for observation and scoring are included in the manual. Scoring is completed after the observation period.

Technical Information

The BRIAAC has not been standardized. However, extensive research has been undertaken for its development and continued improvement. The BRIAAC has content validity in that the scales are "accurate representatives of the various degrees of severity of disturbance" (Manual, p. 72) in autistic and atypical children. Correlations ranging between .43 and .83, obtained by comparing BRIAAC scores and an expert clinician's rating of 20 autistic children, six to 12 years of age, indicated concurrent validity. Factor analysis resulted in varimax loadings ranging from .73 to .90 which confirmed that the scales are tapping a common factor, motivation; however, intercorrelations varied between .54 and .83 confirming the measurement of related behaviors. A study of autistic and retarded children indicated significant mean score differences between the groups in relationship, communication, vocalization, and psychobiological development. A study of normal and autistic children indicated that normal children scored significantly higher than autistic and that their scores were correlated with age.

Rank order correlations based on six to fourteen pairs of observations of 113 autistic children, ranged from .84 to .93. Test-retest reliability estimates were considerably lower.

Effective Use/Comments

Although the BRIAAC is designed to assess low functioning autistic and atypical children in eight areas, the information obtained is insufficient for making screening, performance level, instructional planning, and evaluation of pupil progress decisions. The results may be useful for providing descriptive information about autistic and atypical children who function in these areas below normal four-year-olds.

Reference

Manual

Publisher

Stoelting

Title behavioral Characteristics Age 3 - 18 years
Progression (BCP)

Author Santa Cruz County Office of Time 3 hours - several days
Education

Copyright Date 1973 Type CRT

Intended Purpose of Test

The BCP, designed to serve as a major assessment and instructional tool, provides comprehensive information about student behaviors and development.

Description

The BCP, consisting of 59 checklists of specific behaviors, is composed of 2,400 observable traits or "behavioral characteristics" which have been grouped into the following categories or "behavioral strands":

- | | |
|----------------------------|----------------------------------|
| 1. Health | 22. Listening |
| 2. Attendance/Promptness | (following commands) |
| 3. Feeding/Eating | 23. Adaptive Behaviors |
| 4. Drinking | (reaction to new and routine |
| 5. Toileting | activities) |
| 6. Grooming | 24. Impulse Control |
| 7. Dressing | 25. Interpersonal Relations |
| 8. Undressing | 26. Responsible Behaviors |
| 9. Nasal Hygiene | (obeying rules) |
| 10. Oral Hygiene | 27. Personal Welfare |
| 11. Self-Identification | (safety) |
| (body parts, family) | 28. Self-Confidence |
| 12. Sensory Perception | (class participation) |
| 13. Auditory Perception | 29. Honesty |
| 14. Visual Motor I | 30. Social Speech |
| (fine motor) | 31. Social Eating |
| 15. Visual Motor II | (table manners) |
| (advanced) | 32. Attention Span |
| 16. Gross Motor I - | 33. Reading |
| (coordination) | 34. Task Completion |
| 17. Gross Motor II | 35. Math |
| (sports, fitness) | 36. Practical Math |
| 18. Pre-Articulation | (time, money, measurement) |
| (control of articulation- | 37. Writing |
| related muscles) | 38. Spelling |
| 19. Articulation | 39. Reasoning |
| 20. Language Comprehension | (categorizing, making judgments) |
| (receptive language) | 40. Music and Rhythms |
| 21. Language Development | 41. Arts and Crafts |
| (expressive language) | 42. Pre-Vocational Skills |

- | | |
|--|---|
| 43. Kitchen Skills | 51. Mobility I
(safe movement within the environment) |
| 44. Homemaking Skills | 52. Mobility II
(safe movement to destinations) |
| 45. Outdoor Skills
(garden/house maintenance) | 53. Wheelchair Use |
| 46. Sign Language | 54. Amputation
(parallel bars, rollator, walker, crutches) |
| 47. Fingerspelling | 55. Posture |
| 48. Speechreading
(reads lips) | 56. Swimming |
| 49. Orientation I
(body position, position of objects in the environment) | 57. Articulation I
(vowel & consonant sounds) |
| 50. Orientation II
(estimating and locating destinations, sounds and objects) | 58. Articulation II
(letter sounds & blends) |
| | 59. Health |

The number of strands as well as the specific strands chosen to assess a student should vary with each student; there may be as few as one and as many as 30 per student. However, the greater the number of strands on which a student is assessed, the more complete picture the teacher will have of his or her educational needs.

Test Administration

According to the authors, "the BCP is meant to be an observation tool, not a testing instrument". However, for the purposes of assessment, behaviors may be elicited. The teacher is the principal administrator of the BCP. Although a teacher is capable of conducting the BCP assessment alone, the authors recommend an interdisciplinary setting where several professionals (speech, physical and occupational therapists, school nurses, and school psychologists) can take responsibility for sections of the scales. Under these conditions, administration should take three to eight hours. If administered by a teacher alone, it would take a day or more to test and plan lessons for each student.

Training or experience is necessary for administration.

Technical Information

The BCP is not standardized and does not provide test scores. It was developed by trained professionals who selected items from a wide number of existing scales and tests that relate to developmental skills. The sequencing seems to have been based upon the clinical experience of the authors as well as the norms of those tests from which the items were selected.

No reliability data are provided. Based on review of the manual and the instrument's items, interjudge reliability may be high. The items are clear, behaviorally defined and criterion-referenced.

Although no validation tests were provided, the BCP has high face validity and content validity. The BCP is valid as an observation tool of behavioral characteristics.

Effective Use/Comments

The BCP provides comprehensive information about student behaviors which may be useful for making instructional planning and evaluation of student progress decisions. It is particularly appropriate for developing long- and shortterm objectives on the IEP in the area of adaptive behavior. The information provided is insufficient for making performance level decisions.

The items included in the BCP are simply, clearly and behaviorally defined. If administered properly it can be very helpful in assessing each student's learning objectives. The BCP can effectively be used with all handicapped populations, particularly with the young and severely handicapped.

Since it focuses on adaptive behavior, the BCP may be appropriate for linguistically and culturally diverse students and LEP students provided that language provisions are made for the interviewer.

Reference

BCP Observation Booklet

Publisher

Vort Corporation

Title	<u>Cain-Levine Social Competency Scale</u>	Age	<u>5 - 11 years</u>
Author	<u>L. Cain, S. Levine, & F. Elzey</u>	Time	<u>Untimed</u>
Copyright Date	<u>1963</u>	Type	<u>NRT</u>

Intended Purpose of Test

The Cain-Levine is designed primarily to assess the independence of trainable mentally retarded students.

Description

The Cain-Levine is a rating scale based on information from parents or other individuals very familiar with the student's adaptive behavior. Forty-four items, grouped into four subcategories, are administered in a structured interview. The items measure the habits or typical performance of trainable mentally retarded students in areas of self-help (e.g., dressing, eating), initiative (e.g., self-direction in completing tasks), social skills (e.g., playing, helping others), and communication (e.g., language). The interviewer introduces each item with a general question about the subject's behavior in a particular area and then probes for answers. Thirty-eight items are rated on a 4-point scale, six items on a 5-point scale.

Norm Data

The standardization sample consisted of 716 trainable mentally retarded males and females from California. IQ's of subjects ranged from 25 to 59. Occupational level of the parents or guardians tended to be biased toward an over-representation of lower class families.

Reliability

Odd-even internal consistencies for the total score ranged from .75 to .91. Internal consistencies for individual subtests tended to be lower, ranging from .55 to .95. Test-retest reliability was high (.80).

Validity

Validity was based on item selection and the correlation between social competence and MA and CA. Expert judges were used for item analysis. The scale correlated reasonably well (between .4 and .5) with CA and not very well with IQ (between .09 and .3). Buross (1978) reported that the scale corresponds well with other scales developed for the same general purpose.

Special Alerts/Comments

The Cain-Levine provides an indication of the adaptive behavior of trainable mentally retarded students which may be useful in obtaining supplementary, descriptive information with respect to self-help, initiative, social skills, and communication. As with other third-party

behavior rating scales, the Cain-Levine is not a direct measure of student performance and, consequently, judgment of adaptive behavior requires additional direct behavioral information. With other measures, the Cain-Levine may be useful for making instructional planning decisions, evaluating student progress, and developing short- and longterm objectives on the IEP in the area of adaptive behavior.

Since it focuses on adaptive behavior, the Cain-Levine may be appropriate for trainable mentally retarded students from culturally and linguistically diverse backgrounds and LEP students, provided that language provisions are made for the interviewer.

References

Manual
Buros, 1978
Los Angeles County, 1978
Salvia & Ysseldyke, 1985
Sattler, 1982

Publisher

Consulting Psychologists Press, Inc.

Title	<u>Camelot Behavioral Checklist</u>	Age	<u>All Ages-Mentally Retarded</u>
Author	<u>R.W. Foster</u>	Time	<u>30 minutes</u>
Copyright Date	<u>1974, 1977</u>	Type	<u>CRT</u>

Intended Purpose of Test

The Camelot Behavioral Checklist is designed to identify specific training objectives for mentally retarded individuals of all ages. It provides a summary of those objectives which have been mastered and those which require training.

Description

The Camelot consists of 399 sequentially-listed behavioral descriptions which have been grouped into 40 subdomains and 10 domains: Self-help, Physical Development, Home Duties, Vocational Behavior, Economic Behavior, Independent Travel, Numerical Skills, Communication Skills, Social Behaviors, and Responsibility.

Each item is scored by circling the (+), indicating that the student can do the behavior, or (-), indicating that he or she needs training. Results are reported as raw scores for each domain and total test and are plotted on profile charts.

Test Administration

The Camelot is completed by a trained professional or other staff person, preferably by direct observation. Some information may also be obtained through third-party interviews, although it is not recommended. The ease of administration depends upon the familiarity with the individual being evaluated.

Technical Information

Items were selected from other existing scales. Reliability, based on a study consisting of 50 residents, was above .80, except for Social Behaviors (.69). Total test reliability was .93. Construct validity was developed through an indication of the relationship of scores to other measures of adaptive behavior and general intelligence.

Effective Use/Comments

The Camelot is designed to identify specific training objectives for mentally retarded individuals of all ages which may be useful for making evaluation of students progress decisions and developing shortterm objectives on the IEP in the area of adaptive behavior. Goal setting for profoundly retarded individuals may have to be supplemented through the use of similar instruments (Doucette and Freedman, 1980). Given the lack of information about the standardization sample, the information is insufficient for making performance level decisions. With other measures, results may be useful for making instructional planning decisions.

According to Doucette and Freedman (1980), the "the manual does not specify the population for whom the instrument is intended" (p. 112).

Since it focuses on adaptive behavior, the Camelot Behavioral Checklist may be used with students from culturally and linguistically diverse backgrounds and LEP students provided that language provisions are made for the interviewer.

References

Manual
Doucette & Freedman, 1980

Publisher

Edmark Corporation

Title	<u>Comprehensive Test of Adaptive Behavior (CTAB)</u>	Age	<u>Birth to 60 years</u>
Author	<u>G. L. Adams</u>	Time	<u>Untimed</u>
Copyright Date	<u>1984</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The CTAB is designed to assess the independence of mentally retarded students in six areas: self-help, home living, independent living, social, sensory/motor, and language/academic skills.

Description

The CTAB consists of two forms, one for males, the other for females. There are 527 items for females and 497 for males; these are listed in developmental order within six behavioral categories.

The instrument is completed by an examiner who is familiar with the student. Ratings are based upon the presence or absence of observable or tested behavior and may be verified by using the Parent/Guardian Survey. The Record Form provides space for two test administrations to permit posttesting after instruction.

Norm Data

The author stated that the norm sample for the CTAB may not be representative of the national population of normal or retarded adults. The standardization sample was based on over 6,000 retarded individuals between the ages of early childhood and old age. The sample was stratified by age, sex, educational setting, "school", "non-school", region, and severity of handicapping condition. Subjects from the "non-school" setting were drawn from 15 states including Florida while those from the "school" setting came exclusively from Florida.

Reliability

Split-half reliability ranged above .86 with most correlations over .95. Test-retest estimates ranged above .81. Interrater reliability was high.

Validity

Criterion-related validity was demonstrated by a higher correlation with the VSMS (.68) and lower correlations with the WISC-R (.55) and Stanford-Binet (.38). Construct validity was developed through an indication of the relationships of similar and dissimilar domains. Content validity was indicated through consistency with current definitions of adaptive behavior. Predictive validity based upon comparisons between the NABC and CTAB was high.

Special Alerts/Comments

The CTAB provides an indication of a student's level of independence which may be helpful in obtaining descriptive information with respect to self-help, home living, independent living, social, sensory/motor, and language/academic skills. The information provided is sufficient for making instructional planning and evaluation of progress decisions. As with other third party scales, CTAB is not a direct measure of student performance and, consequently, judgment of social competence will require direct behavioral information. When administered with other measures, the CTAB may be useful for determining level of performance.

Test administration may be time-consuming but scoring is relatively simple.

The CTAB may also be categorized as a criterion referenced test since it provides information on the extent to which a student possesses specific skills in specific areas. Task analyses and long- and shortterm objectives may be derived for the IEP in the area of adaptive behavior.

Based upon the limited standardization sample and the author's admitted need to update the performance rankings, interpretations from this test should be used cautiously with students in New York City.

Since it focuses on adaptive behavior, the CTAB may be appropriate for students from culturally and linguistically diverse backgrounds and LEP students provided that language provisions are made for the Parent/Guardian Survey and for the Language/Academic category.

Reference

Manual

Publisher

The Psychological Corp.

Title	<u>Normative Adaptive Behavior Checklist (NABC)</u>	Age	<u>Birth to 60 years</u>
Author	<u>G. L. Adams</u>	Time	<u>Untimed</u>
Copyright Date	<u>1984</u>	Type	<u>NRT</u>

Intended Purpose of Test

The NABC is designed to assess the independence of mentally retarded students in six areas: self-help, home living, independent living, social, sensory/motor, and language/academic skills.

Description

The NABC is the short form of the CTAB. It contains 120 items which identify adaptive behavior deficits for ages birth to 21 years. The NABC is intended primarily as a screening test.

Norm Data

The standardization sample was based on over 6,000 nonretarded individuals between the ages of birth and 21. The sample was stratified by age, sex, region, and race and was consistent with the 1980 U.S. Census. Results of the test are reported as percentile ranks.

Reliability

Intrarater reliability indicated a strong agreement between the rater and items duplicated in different locations. Additional reliability information was unavailable.

Validity

Construct validity was developed through an indication of the relationships of similar and dissimilar domains. Content validity was indicated through consistency with current definitions of adaptive behavior. Predictive validity based upon comparisons between the NABC and CTAB was high.

Special Alerts/Comments

The NABC provides an indication of a student's level of independent behavior which may be useful for making screening decisions. As with other third party scales, NABC is not a direct measure of student performance and, consequently, judgment of social competence will require direct behavioral information. The information provided is insufficient for making performance level, instructional planning, and evaluation of progress decisions.

Based upon the limited norm sample and the author's admitted need to update the performance rankings, interpretations from this test should be used cautiously with students in New York City.

Since it focuses on adaptive behavior, the NAEC may be appropriate for students from culturally and linguistically diverse backgrounds and LEP students provided that language provisions are made for the Parent/Guardian Survey and for the Language/Academic category.

Reference

Publisher

Manual

The Psychological Corp.

Title	<u>Personal/Social Behavior Curriculum Guide (PSBCG)</u>	Age	<u>3 - 14 years</u>
Author	<u>Southern Westchester BOCES</u>	Time	<u>Untimed</u>
Copyright Date	<u>1978</u>	Type	<u>CRT</u>

Intended Purpose of Test

The rating scale of the PSBCG is intended to evaluate a student's personal and social behaviors with respect to himself/herself, his/her peers, and his/her teachers.

Description

The rating scale is divided into four subtests:

- | | |
|----------------------------------|--|
| I. Task - Individual | a. Reactions to being assigned a task |
| | b. Behavior while working on the task |
| | c. Behavior at the end of the time assigned for the task |
| II. Social Interaction | Relationship to peers |
| III. Relationship to the Teacher | Relationship to teacher |
| IV. Emotional Responses | a. Warmth of emotions |
| | b. Behavior under stress |
| | c. Symptomatic behavior |

The student is rated as always, often, sometimes, rarely, and never behaving in the manner described. The clusters of skills are categorized by age levels. The level of mastery is directly related to both the objective and the age level involved.

Test Administration

The PSBCG is completed entirely by the teacher, guidance counselor, or other professional. No special training or materials are required. No information is provided regarding the length of time to be spent observing the student before completing the scale.

Technical Information

The PSBCG is part of a curriculum created by a committee composed of representatives from eleven consortium school districts in Southern

Westchester, Southern Westchester BOCES staff members, and N.Y.S. Title IV-C staff members to provide a comprehensive series of annual goals and instructional objectives in the affective area. It is a result of the New York State Education Mandate, Chapter 853. Evaluation items were based on those included in the September 1978 revision of the Southern Westchester BOCES Affective Rating Scale. No information is provided with respect to reliability and validity.

Effective Use/Comments

The PSBCG is intended to evaluate a student's personal and social behaviors with respect to himself/herself, his/her peers, and his/her teachers. The information obtained is sufficient for developing shortterm objectives on the IEP in the area of adaptive behavior. With other measures, the findings may be useful for making performance level, instructional planning, and evaluation of student progress decisions.

In addition to specifying objectives, the manual provides specific activities regarding instruction and curriculum development.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

Reference

Manual

Publisher

Board of Cooperative Educational
Services (BOCES)-
Southern Westchester County

Title	<u>Preverbal Assessment-Intervention</u>	Age	<u>Birth - Adult</u>
	<u>Profile (PAIP)</u>		<u>(severely/profoundly handicapped)</u>
Author	<u>P. Connard</u>	Time	<u>Untimed</u>
Copyright Date	<u>1984</u>	Type	<u>CRT</u>

Intended Purpose of Test

The PAIP was designed to assess the sensory, motor, and communicative skills of severely, profoundly, and/or multihandicapped individuals with a functioning level between zero and eight months.

Description

The PAIP consists of a series of tests which are divided into three stages that follow Piaget's theory and framework of infant development. Stage I, developmental ages zero to one month, is subdivided into five tests; Stage II, developmental ages one to four months, is subdivided into four tests; Stage III, developmental ages four to eight months, is subdivided into three tests.

The tests, Preliminary Placement Evaluations, Diagnostic Profiles, and Narrative Plans are included in a 31-page record booklet. The manual provides detailed directions for administration and scoring and a Glossary.

Test Administration

The PAIP is individually administered by a trained practitioner who indicates observed, reported, or elicited behaviors which occur in the natural environment. Results are recorded on grids which form a profile of the individual's pattern of functioning. Interpretation of the profile assists in the identification of instructional target behaviors and the development of intervention programs.

Technical Information

Inter-observer reliability was high with agreement percentages of 88% or higher. The sample consisted of 50 severely, profoundly, and/or multihandicapped individuals who were 8 to 37 years old and who functioned at an infant level of behavior. Test-retest reliability was indicated by the consistency of Stage placement for 15 individuals. Concurrent validity was demonstrated by the PAIP's relationship to the Early Learning Accomplishment Profile (ELAP).

Effective Use/Comments

The PAIP was designed to assess the sensory, motor, and communicative skills of severely, profoundly, and/or multihandicapped individuals with a functioning level between zero and eight months. With other measures, the findings may be useful for making screening, performance level, instructional planning, and evaluation of student progress decisions. The information is sufficient for developing shortterm objectives on the IEP in the area of adaptive behavior.

As with other third party scales, the PAIP is not a direct measure of student performance and, consequently, judgment of adaptive competence will require direct behavioral information.

Due to the complex, technical, and lengthy administration, recording, and scoring, the PAIP should be administered by a highly trained practitioner.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

Reference

Manual

Publisher

A.S.I.E.P. Education Co.

Title	<u>Scales of Independent Behavior (SIB)</u>	Age	<u>Birth - 40+ years</u>
Author	<u>R. H. Bruininks, R. W. Woodcock, R.F. Weatherman, & B.K. Hill</u>	Time	<u>Untimed (10-60 minutes)</u>
Copyright Date	<u>1984</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The SIB is designed to measure functional independence in motor development, social development, language, self-help, and community adaptation for handicapped and nonhandicapped infants, children, and adults.

Description

The SIB is a set of tests which are individually administered through a structured interview with a respondent who is familiar with the subject. The interviewer presents items orally while the respondent looks at the items and the scoring key on a flip-book or easel format.

The Early Development Form is designed to assess the adaptive development of children from birth to three years of age. The Full Scale and Short forms may be used from infancy to age 40 and over.

The Full Scale Form contains 226 task items in 14 subscales organized into four domains which are rated according to a 4-point scale. The Early Development and Short Forms contain 32 items and require a shorter administration time than the Full Scale. Maladaptive behavior is measured by the Problem Behaviors Scale, which includes eight major categories.

Norm Data

The standardization sample was based on approximately 1,800 subjects between the ages of infancy to adult, which were randomly drawn from over 40 communities. The sample was stratified by sex, race, Spanish origin, socioeconomic status, geographic region, and community size according to the 1980 U.S. Census. Additional data were obtained on over 1,000 handicapped and nonhandicapped people including mentally retarded, learning disabled, behavior disordered, and hearing impaired subjects.

Test results are reported in terms of age equivalents, percentile ranks, standard scores, normal curve equivalents (NCEs), relative performance indexes (RPIs), and maladaptive behavior indexes (MBIs).

Reliability

Split-half reliabilities for individual subscales, based on handicapped samples, calculated across all age levels, ranged from .73 to .86, while those for the four domains ranged from .83 to .93. The internal consistency of the Broad Independence Scale was .96.

High internal consistency was reported with handicapped samples (.88 to .95 for individual subscales, high .90's for domains). Test-retest reliability estimates, available only on Short Form domains, based upon 67 students retested at an interval of one to four weeks, ranged from .71 to .96. Interrater reliability was high.

Validity

Content validity was indicated through consistency with current theories of personal and community adjustment including both adaptive and maladaptive areas. Construct validity was developed through an indication of the relationship of scores to ages, handicapping condition, and behavioral development for the standardization sample. Concurrent validity was demonstrated by the SIB's relationship to other adaptive and problem behavior measures.

Special Alerts/Comments

The SIB provides information about personal and community independence with respect to motor development, social development, language, self-help, and community adaptation of handicapped and non-handicapped infants, children, and adults. With other measures, the findings may be useful for screening, program evaluation, making instructional planning decisions, and determining level of performance. The information is useful developing shortterm objectives on the IEP in the area of adaptive behavior. As with other third party social scales, the SIB is not a direct measure of student performance and, consequently, judgment of social competence will require direct behavioral information.

Scoring procedures are lengthy and may be somewhat confusing. Since the response booklet lists item numbers rather than tasks, IEP information must be obtained by comparing responses with the descriptions in the flip-book.

Since it focuses on adaptive behavior, the SIB may be appropriate for students from culturally and linguistically diverse backgrounds and LEP students provided that language provisions are made by the interviewer.

Reference

Manual

Publisher

DLM Teaching Resources

Title	<u>Tarc Assessment Inventory for Severely Handicapped Children (TARC)</u>	Age	<u>3 - 16 years</u>
Author	<u>W. Sailor & B.J. Mix</u>	Time	<u>10 minutes</u>
Copyright Date	<u>1975</u>	Type	<u>CRT</u>

Intended Purpose of Test

The TARC is designed to assess the abilities of retarded or severely handicapped students with regard to adaptive skills related to education.

Description

The TARC consists of a manual and a booklet, including directions for scoring, a score sheet, and cover and profile sheets for standard scores.

The TARC includes the following four subtests:

<u>Subtest</u>	<u>Skills Assessed</u>
Self-Help	Toileting, washing, eating, clothing
Motor	Small muscle, gross motor, pre-academic (specific motor skills of direct relevance for classroom education)
Communication	Expressive, receptive, pre-academic (classroom-related specific communication skills)
Social Skills	Following directions, group behavior, individual behavior, interaction with peers and with adults, cooperation and emotional control)
	Pre-academic (Specific classroom-related social skills)

Test Administration

The rater should be familiar with the student before using the inventory; the manual recommends informal observations for a minimum of three weeks. The rater then assesses the student on two types of items, scaled and categorical, both of which describe current ability. Item scores are added together for subsection totals; subsection totals are added up for section totals; section totals are then added together to become the total raw score. Results are recorded on a profile sheet.

Technical Information

The inventory was standardized on a sample of 283 severely handicapped students of both sexes, ranging in age from 3 to 16. The majority of the sample would be considered trainable mentally retarded.

Interrater reliability ranged from .59 to .78 for different skill areas, and .85 for the total test. Test-retest reliability was .80.

Effective Use/Comments

The TARC was designed to assess the abilities of retarded or severely handicapped students with regard to adaptive skills related to education. With other measures, the findings may be useful for making performance level, instructional planning, and evaluation of student progress decisions. The information is sufficient for developing shortterm objectives on the IEP in the area of adaptive behavior.

The interrater reliability level suggests that caution be exercised if anyone other than the rater uses the instrument to obtain information about a specific student. "Its use as a screening tool with unfamiliar clients is not recommended" (Doucette & Freedman, 1980, p. 215). In addition, too little information was provided concerning the standardization sample to use the scaled scores as a measure of ability.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

References

Manual
Doucette & Freedman, 1980

Publisher

Edmark Corp.

Title	<u>Vineland Adaptive Behavior Scales (VABS)</u>	Age	<u>Birth to 18-11 years</u>
Author	<u>P.L. Harrison</u>	Time	<u>20-90 minutes</u>
Copyright Date	<u>1985</u>	Type	<u>NRT, CRT</u>

Intenueu Purpose of Test

The VABS is designed to assess personal and social sufficiency in four domains: Communication, Daily Living Skills, Socialization, and Motor Skills.

Description

The 1985 VABS is a revision of the 1965 Vineland Social Maturity Scale (VSMS). Three versions, each of which measures adaptive behavior in four domains and subdomains, are available: The Interview Edition, Survey Form; The Interview Edition, Expanded Form; and the Classroom Edition. Each form differs in the number of items and materials and the method of administration. The Survey Form provides general assessment; the Expanded Form offers a more systematic basis for preparing individual educational programs; the Classroom Edition provides assessment of classroom behavior. The Survey and Expanded Forms are administered by a trained interviewer to a parent or guardian; the Classroom Edition is administered to a teacher. Ratings are made based upon the presence, partial presence, or absence of the behavior relevant to the scale items.

The Survey Form contains 297 items which are listed in developmental order within five general domains and subdomains. The Expanded Form contains 577 items, including all items from the Survey Form, which are listed in developmental order within clusters. The Classroom Edition contains 244 items, including some from both the Survey and Expanded Forms and additional academically related items. The trained teacher assigns scores to items, grouped by subdomains within domains, and identifies the scores as Observed or Estimated Performance.

Scores from the four adaptive domains are combined to form the Adaptive Behavior Composite.

A fifth domain, The Maladaptive Behavior Domain, is included in the Survey and Expanded Forms for optional use. Frequency of behavior as well as severity of behavior are assessed. The Maladaptive Behavior domain is administered to individuals five years of age or older. The parent or caregiver is asked to indicate whether certain maladaptive behaviors occur usually, sometimes, or never.

Norm Data

The standardization sample was tested between September 1981 and May 1982. The sample was stratified by age, sex, ethnic group, community size, geographic region, and parental educational level and corresponded to 1980

U.S. Census figures. The Survey and Expanded Forms were standardized on 3,000 individuals from birth to 18-11 years. The Classroom Edition was standardized on 3,000 students from 3 to 12-11 years.

Raw scores are converted to derived scores to make domain comparisons possible. Scores are reported as percentiles, adaptive levels, age equivalents, and stanines for each subdomain and domain; an Adaptive Behavior Composite may also be obtained. Supplementary norms of mentally retarded, emotionally disturbed, visually handicapped, and hearing-impaired individuals permit comparisons of students with others with the same handicapping condition. Both norm referenced and criterion referenced information available on all forms.

Reliability

Internal consistency for domains on all forms generally ranged above .83; coefficients for subdomains ranged above .69. Test-retest and interrater reliability estimates, available only on the Survey Form, were generally lower than that reported for internal consistency. Reliabilities for the Interview Forms are fairly high; those for the Classroom Edition "do not appear adequate" (Salvia & Ysseldyke, 1985, p. 405).

Validity

Construct validity was developed through an indication of the relationship of scores to age for the standardization sample and to handicapping condition for the supplementary normative group. Content validity was developed through the author's definition of adaptive behavior. Criterion-related validity was indicated by comparisons with several tests, such as the original VSMS, ABIC, Kaufman Assessment Battery for Children (K-ABC), Peabody Picture Vocabulary Test (PPVT), and ABS-SE. Results with the K-ABC and PPVT revealed the expected low correlations and supported the assumption that an adaptive behavior scale and an intelligence and achievement scale measure different areas of functioning. Correlations between the ABIC, the ABS-SE and the VABS ranged from .18 to .92.

Special Alerts/Comments

The VABS provides an indication of a student's level of personal and social sufficiency which may be helpful in obtaining descriptive information with respect to communication, daily living skills, socialization, and motor skills. As with other third party scales, the VABS is not a direct measure of student performance and, consequently, judgment of social competence will require additional information obtained through direct behavioral observation.

The information is sufficient for developing short- and longterm objectives on the IEP in the area of adaptive behavior. With other measures, the results may be useful for making screening, performance level, instructional planning, and evaluation of student progress decisions.

The author stated that each domain on the Survey and Expanded Forms should be administered in its entirety and all domains should follow their standardized order, even if some domains are omitted. Shortterm program planning information can be obtained from the Expanded and Classroom Forms.

In contrast to the VSMS, the VABS reflects an increased number of test items, a choice of three forms of administration procedures, and standardization with nonhandicapped as well as handicapped individuals. Furthermore, ratings on the VABS are determined by partial presence of behavior, not only total presence or absence as with the VSMS. Scoring procedures are detailed and complex; software scoring materials are available.

The VABS may be appropriate for students from culturally and linguistically diverse backgrounds and LEP students provided that the language provisions are made by the interviewer. Caution should be observed in making individual decisions because of low reliability estimates and limitations of the standardization sample (Salvia & Ysseldyke, 1985).

References

Manual
Salvia & Ysseldyke, 1985

Publisher

American Guidance Service Inc.

Title	<u>Weller-Strawser Scales of Adaptive Behavior for the Learning Disabled (WSSAB)</u>	Age	<u>6 - 18 years</u>
Author	<u>C. Weller & S. Strawser</u>	Time	<u>Untimed</u>
Copyright Date	<u>1981</u>	Type	<u>CRT</u>

Intended Purpose of Test

The WSSAB is intended to measure the adaptive behavior of learning disabled elementary and secondary students in social coping, relationships, pragmatic language, and production.

Description

The WSSAB is available for two levels: (1) elementary for use with learning disabled students ages six through twelve, and (2) secondary for use with students ages thirteen through eighteen.

Each scale is composed of four subtests consisting of Social Coping, Relationships, Pragmatic Language, and Production. The items in each subtest consist of paired descriptions of the range of a specific adaptive behavioral characteristic.

<u>Subtest</u>	<u>Contents</u>
1. Social Coping	<u>Elementary:</u> use of time and space; effect of change; handling worries or depression; handling criticism and problem solving <u>Secondary:</u> dress; following site directions; orientation in space; acceptance of change and disorganization; understanding cause and effect relationships
2. Relationships	<u>Elementary:</u> with peers and authority figures; manner of seeking amusement; expectations of others' behavior; implications of possessiveness in relationships <u>Secondary:</u> expectation and acceptance of others' behavior; types of relationships chosen; use of expletives and profanity; degree of involvement in social activities; social awareness
3. Pragmatic Language	<u>Elementary:</u> use of verbal language; use of non-verbal language; need for verbalization; use of gestures and humor; ability to use language to cover up learning deficits

	<u>Secondary:</u>	understanding voice inflexion and practical jokes; relationship between gestures, nonverbal communication, or body language to communication; interpretation of others' feelings about student; appropriateness and description of responses; use of slang
4. Production	<u>Elementary:</u>	how tasks are approached; how past learning is transferred to new skills; how working time is spent; how student responds to various activities
	<u>Secondary:</u>	relationship between group size and production; ability to shift tasks; organization of work space; need for stimulation to continue working; desire to participate in tasks and task strategies

Test Administration

The authors recommended completion of the WSSAB as early as possible in a student's entrance into a learning disabilities program, given a reasonable period for observation.

Scoring procedures are extremely simple. Space is provided on the first page of each record form for the development of a profile. The scales need not be completed in one sitting.

Technical Information

Items were selected from tape recorded interviews with teachers who taught learning disabled students. Five experts in the field of learning disabilities read the items for clarity, appropriateness of language, and continuity between choices. Field-testing was conducted by teachers of learning disabled students from urban and rural areas in six states, none of which were located in the Northeast. A total of 236 students, ages six to eighteen, formed the sample.

Internal consistency estimates for subtests were reported to be greater than .94 for either scale. Interrater reliability coefficients were .88 and .89 for the Elementary and Secondary Scales, respectively.

The authors maintained that content validity was established through the methods of item development. Interitem correlations resulted in coefficients of .48 to .78 for the Elementary Scale and .53 to .86 for the Secondary Scale.

Effective Use/Comments

The WSSAB measures the adaptive behavior of learning disabled students in social coping, relationships, pragmatic language, and production. As with other third party scales, the WSSAB is not a direct measure of student performance and consequently judgment of adaptive competence will require direct behavioral information. With other measures, the findings may be useful for making instructional planning and evaluation of student progress decisions. The information is sufficient for developing shortterm objectives on the IEP in the area of adaptative behavior.

In addition to specifying objectives, the manual provides strategies for instruction. The authors stated that the scales are to be used only after a student has been diagnosed as learning disabled; it is not intended for use with students diagnosed as having any other handicapping condition or as having no handicapping condition.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

Reference

Publisher

Manual

Academic Therapy Publications .

COGNITIVE FUNCTIONING

Measures of cognitive functioning assess general and specific learning aptitude and conceptual reasoning based upon past experience; i.e., learning. The purpose of these measures is generally to predict future performance of the individual, typically in the school setting.

Thus, some psychometricians have labeled such tests as "scholastic aptitude tests" (Thorndike & Hagen, 1977). Cognitive tests do not measure innate potential. These measures, frequently called intelligence tests, differ from achievement tests only in "being less dependent upon specific segments of school instruction and in being used to make inferences about the future, rather than to draw conclusion referring to the past" (Thorndike & Hagen, 1977, p. 297).

Only a few cognitive instruments include adequate representation of minority populations in their standardization samples. The examiner should be familiar with the literature that addresses this issue and weigh this factor in selecting a test for a particular student.

Salvia and Ysseldyke (1985) emphasized that it is not only important to know what behavior the test samples but also how "acculturation" affects test performance. Therefore, results with culturally and linguistically diverse students should be interpreted cautiously.

Title	<u>Arthur Adaptation of the Leiter Performance Scale (AALIPS)</u>	Age	<u>3-8 years</u>
		Time	<u>Untimed</u>
Author	<u>G. Arthur</u>	Type	<u>NRT</u>
Copyright Date	<u>1950</u>		

Intended Purpose of Test

The AALIPS is designed to assess the intelligence of students who have difficulty responding verbally.

Description

An individually administered test, the test materials are identical to the Leiter International Performance Scale through the 12-year-level. The material consists of response blocks which correspond to stimulus cards. The test is administered by presenting a stimulus card and pantomiming directions. The student is required to organize the blocks to correspond to the stimulus cards. The manual includes no key to correct answers to test questions. Both an MA and IQ can be computed.

Norm Data

Norm data are not included in the manual. However, the test was standardized on 289 homogeneous middle-class, mid-western, metropolitan students. Few of the students in the standardization sample had difficulty responding verbally.

Reliability

No reliability data were reported in the manual. However, Anastasi (1976) has reported split-half reliability of .90. Test-retest reliability was reported as good after a 3-year interval.

Validity

Arthur reported evidence of concurrent validity. Correlations with the S-8 for 4- to 8-year-olds ranged from .69 to .93 for mentally retarded, between .56 to .86 for brain injured; it correlated more highly with the Performance Scale (.79 to .80) than with the Verbal Scale (.40 to .78) of the WISC.

Special Alerts/Comments

The APPLIPS provides information about the intellectual performance of students who have difficulty responding verbally which is insufficient for determining a level of performance. The severe technical inadequacies of this test are such that the results of the AALIPS may be used only to obtain descriptive information about intellectual/cognitive ability.

According to Salvia and Ysseldyke (1981), "its use should be restricted to procurement of qualitative information by only the most experienced examiners" (p. 183). As an informal tool, the most experienced examiner may obtain useful information with severely impaired students.

Because of its limited standardization and previously cited technical inadequacies, results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

references

Anastasi, 1976
Los Angeles Guide, 1978
Salvia & Ysseldyke, 1985
Sattler, 1982

Publisher

Stoelting Company

Title	<u>Assessment of Basic Competencies (ABC)</u>	Grades	<u>Pre-K-9</u>
Author	<u>J. P. Somwaru</u>	Time	<u>2-3 Hours</u>
Copyright Data	<u>1981</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The ABC is an individually administered measure of cognitive development, providing information on student competencies or abilities and their nature. The test is useful for obtaining information on students in grades Pre-K to 9 for whom such information is needed because of problems in learning.

Description

The ABC consists of eleven subtests organized into the three domains of information processing, language, and mathematics. Information processing includes the fundamental strategies for learning; subtests consist of observing skills including visual-motor processes and visual memory, organizing skills including grouping or classifying and ordering, and relating skills including analyses and making inferences. Language subtests consist of understanding words, comprehending expressions which focus on understanding grammatical structures, producing expressions, decoding, and reading comprehension skills. Mathematics subtests consist of knowing number and operations, understanding concepts, and solving problems.

Administration takes two to three hours in two to four sessions plus 1/2 hour for scoring. If used as a norm referenced test, the ABC yields developmental ages or percentiles; as a criterion referenced test, results indicate achievement of skills.

Norms

Test development was based on the latent trait model which is regarded as a positive development in testing. The ABC was normed on less than 2000 students per trait or subtest between grades Pre-K to 9. Approximately eight percent were in the northeast, represented by Maine, New Jersey, and New York, where students in White Plains and districts 24 and 28 participated. Some 15% were black and other. It is apparent, that with 49% of the sample from the north central and almost 50% from rural and small city areas, this test is representative of "middle America".

Reliability

Reliability was based on small samples indicating generally adequate test-retest correlations of equivalence for each of the 11 traits or cluster. Item reliabilities were not presented.

Validity

Validity was indicated in three ways: (1) Guttman's facet theory was used to show construct validity; however, intercorrelations of clusters did not provide strong support for the theory. (2) Content validity was based upon test development and the opinion of experts. (3) Criterion validity based on small N's indicated adequate correlation with SRA in grades 2 and 6, high correlation with teacher ratings in grade 6, and high correlation with age, grade, and WISC age.

Special Alerts/Comments

The ABC is designed to measure student competencies in cognitive development which may be useful for screening decisions. Given the limitations of the standardization sample, the results of the ABC are insufficient for determining performance level. The results may be used only to obtain descriptive information about intellectual/cognitive ability.

The ABC may be administered by a psychologist, educational evaluator, or teacher who has been trained to give the test. It has complex administrative characteristics. In addition to basal-ceiling determinations, the test administrator must "juggle" materials, scrap paper, pencil, and test booklets with the student. Recording scores is relatively simple, although 30 minutes may be required for scoring.

The ABC should be used cautiously in New York City until an empirical study is undertaken to compare scores with those of other known instruments, such as the WISC-R. The usefulness of the test with all students in New York City, not only LEP students and students with culturally and linguistically diverse backgrounds, is unsubstantiated and caution should be exercised when determining the intellectual/cognitive level of such students.

Reference

Manual

Publisher

Scholastic Testing Service, Inc.

Title	<u>Bayley Scales of Infant Development (BSID)</u>	Age	<u>2 month-2 1/2 years</u>
Author	<u>N. Bayley</u>	Time	<u>Untimed</u>
Copyright Date	<u>1969</u>	Type	<u>NAT</u>

Intended Purpose of Test

The BSID is designed to assess the developmental status of young children's mental and motoric abilities.

Description

The BSID is comprised of three components: the Mental Scale, the Motor Scale, and the Infant Behavior Record (IBR). The Mental Scale assesses: the ability of the infant to respond to sensory-perceptual stimuli; the early acquisition of memory and problem-solving ability; the beginnings of verbal communication and early vocalization; and early indications of abstract thinking.

The Motor Scale assesses: body control and fine and gross motor skills. The Infant Behavior Record (IBR) is used to evaluate the child's social and objective orientation toward his/her environment through notations of the child's interests, emotions, attitudes, energy, activity, and approach or withdrawal from stimulation. The IBR is filled in after the Mental and Motor Scales have been completed and the child has left the examining room.

Norm Data

The standardization sample consisted of over 1,200 children. The sample was controlled for sex, ethnicity, urban-rural residence, and education of the head of the household within each age group. Institutionalized children with severe emotional or behavioral problems and children born one month prematurely were excluded; children from bilingual homes one year or older who showed difficulty in using English were also excluded from the sample.

Reliability

Split-half reliability coefficients for the Mental Scale ranged from .81 to .93 with a median value of .88. Coefficients for the Motor Scale tended to be somewhat lower (.68 to .92, median .84). Standard errors of measurement are included in the manual, thereby making test score reporting more precise.

Validity

No data on validity were reported, but many scales and checklists are derived from the BSID.

Special Alerts/Comments

The BSID provides a measure of the developmental status of young children's mental and motoric abilities which may be useful in determining a level of cognitive functioning of young children.

The examiner should be well trained in the application of the Bayley Scales as well as be familiar with normal and deviant development.

A potential drawback to using the BSID is the necessity for obtaining certain materials not included in the test kit; e.g., a standard set of stairs, a standard walking board, dimensions of which are delineated and of critical importance.

It should be noted that infant tests fail to predict later IQs.

Because of the limited standardization sample, the BSID should be interpreted cautiously with children of diverse cultural and linguistic backgrounds. Scores may not be reported for bilingual and/or LEP students.

References

Manual
Buros, 1978

Publisher

The Psychological Corp.

Title	<u>Blind Learning Aptitude Test (BLAT)</u>	Age	<u>6 - 16 years</u>
Author	<u>T.E. Newland</u>	Time	<u>20-45 minutes</u>
Copyright Date	<u>1977</u>	Type	<u>NRT</u>

Intended Purpose of Test

The BLAT is designed to assess the learning aptitude of young blind students.

Description

This is a non-verbal individually administered aptitude test. Items are drawn in large part from tests developed to avoid cultural bias, e.g., The Culture Free Intelligence Test, Raven's Progressive Matrices. Each test and training item is presented on a separate plastic sheet held in a three-ring loose-leaf binder. Items are constructed of dots and lines presented in bas-relief. No knowledge of braille is required. There are 49 test items, 12 training items, and a "test age" or "learning aptitude age".

The BLAT yields a "learning quotient" with a mean of 100 and an SD of 15. There is little guidance in interpreting an IQ. According to the manual, the test is more discriminating at the younger ages (6 through 12).

Norm Data

The BLAT was standardized on a sample of 961 residential and day-school functionally blind students in twelve states (two West Coast, two East Coast, five Mid-Western, three Southeastern), socioeconomically and racially representative of the U.S. (Johnson, 1976).

Reliability

Internal consistency was .93 for the total sample. A test-retest correlation coefficient, based on a 7-month interval, was reported to be .87 for a sample of 93 students, 10 to 16 years of age.

Validity

Concurrent validity was based on correlations with the Hayes-Binet and WISC-verbal, .74 and .71, respectively.

Special Alerts/Comments

The BLAT provides a measure of learning aptitude of young blind students which may be useful for determining the student's intellectual/cognitive ability. While its validity needs further study, the "BLAT is currently the most adequate test for assessing learning aptitude of young blind children" (Salvia & Ysseldyke, 1985, p. 181).

Given the limitations of the standardization sample, use of the BLAT with culturally and linguistically diverse blind students and blind LEP students has not been established. Scores should not be reported for bilingual and/or LEP blind students.

References

Manual
Johnson, 1976
Saivla & Ysselayke, 1985

Puolisner

University of Illinois Press

Title	<u>Boehm Test Basic Concepts - Revised</u> <u>(BTBC-R)</u>	Grade	<u>K - 2</u>
Author	<u>A. Boehm</u>	Time	<u>Untimed</u> <u>(15-20 minutes)</u>
Copyright Date	<u>1986</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The BTBC-R is designed to assess a student's mastery of concepts essential for early school achievement. It is intended to identify students with basic concept deficiencies.

Description

Two alternate forms, Form C and Form D, are available. Each form consists of 50 pictorial items arranged according to order of increasing difficulty. The 50 items of each form are divided into two booklets of 25 test items each. An Applications booklet can be used with either form, or by itself. The Application booklet assesses mastery of concepts used in combination, used in sequences, and used to make comparisons.

The BTBC-R is group administered. The classroom teacher reads aloud the directions. The students are instructed to mark the picture which best answers the question read aloud by the teacher. It is important to follow the detailed directions of the BTBC-R exactly.

A Class Record Form is used as the scoring key as well as being an aid for the teacher to gain an understanding of the reasons underlying test errors.

Norm Data

National norms were established for the BTBC-R at both the beginning and end of the school year. School district size and geographic region were taken into account using 1981 national statistics published by the National Center for Educational Statistics. Standardization occurred in 1983-84. The full range of socioeconomic status in the U. S. was represented. The norm samples used for Form C and D were 4,604 and 6,923, respectively. Form C and D were found to be equivalent, based on a sample of 199.

Reliability

Split-half reliability coefficients and standard errors of measurement were computed for students in the spring 1983 standardization sample. Separate data are available for each socioeconomic level within each of the three grades tested. Reliability coefficients ranged from .55 to .87.

The median coefficient was .77. For the Application booklet, coefficients ranged from .72 to .82 with a median of .76 for grades 1 and 2. Standard errors of measurement were reported; the results indicated a satisfactory level of precision.

Content validity was determined by factor analytic studies. The BTBC-R is best described as a test of verbal concept acquisition.

Validity

The author reported a "reasonable strong relationship between BTBC-R scores and achievement measures gathered one year later." (Manual, p. 59). Seventeen correlation coefficients ranged from .28 to .64 with a median of .44.

Special Alerts/Comments

The BTBC provides a measure of a student's mastery of concepts. The information provided is not sufficient for making performance level decisions. The information may be useful for making evaluation of student progress and instructional planning decisions and for developing shortterm objectives on the IEP.

Since each of the 50 concepts on Form C and D is measured by only one item, the author correctly stated that the BTBC-R "cannot be considered a criterion referenced measure for individual concepts where three to five items per concept would be desirable" (Manual, p.2). However, a criterion number of concepts can be established by the teacher as an objective; e.g., "by the end of the semester, the student will properly identify 48 of 50 fundamental concepts."

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously. A score may not be reported for bilingual and/or LEP students.

Reference

Manual

Publisher

The Psychological Corp.

Title Bracken Basic Concept Scale (BBCS) Age 2 - 8 years
Author B. A. Bracken Time Untimed
Copyright Date 1984 Type NRT, CRT

Intended Purpose of Test

The BBCS is intended to assess a child's conceptual knowledge.

Description

The BBCS consists of two separate instruments: two alternate form screening tests and a diagnostic scale which consists of an easel-type stimulus manual, an examiners manual, and a record form. The screening tests are primarily for kindergarteners and first graders to identify those children who may benefit from more intensive assessment. The BBCS can be administered by teachers, evaluators, or paraprofessionals. The BBCS screening tests can be administered either in groups or individually. The BBCS Diagnostic Scale is individually administered.

Over 250 concepts are assessed on the Diagnostic Scale, which are divided into eleven categories:

Color	Social/Emotional
Letter Identification	Size
Numbers/Counting	Texture Material
Comparisons	Quantity
Shapes	Time/Sequence
Direction/Position	

Response to the BBCS can be verbal or gestural (indicated by pointing). Aphasic or mute children may, therefore, be assessed with this instrument.

Norm Data

The Screening Tests of the BBCS were normed on over 800 kindergarten and first-grade children. The Diagnostic Scale was standardized on 1,100 children. The standardization sample was stratified on age, sex, ethnic group, geographic region, community size, and socioeconomic status based upon the 1980 U.S. Census.

Standard scores, concept ages, and percentile ranks can be generated from results of the BBCS.

Reliability

Test-retest reliability, determined by retesting 27 children over a 2-week interval, resulted in a coefficient of .97. However, caution should be exercised in interpreting this result since the sample used was so small.

Validity

Validity studies compared the BBCS Diagnostic Scale with measures of receptive language, preschool readiness, and knowledge of basic concepts. Coefficients ranged from .68 to .88.

Special Alerts/Comments

The BBCS provides an indication of a child's conceptual knowledge which may be useful for making screening decisions. The Diagnostic Scale also provides criterion referenced information useful for making instructional planning decisions for young children and developing long- and shortterm objectives on the IEP in the area of concepts. Given the lack of technical information the BBCS is insufficient for determining intellectual/cognitive functioning.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously. Scores should not be reported for bilingual and/or LEP students.

Reference

Manual

Publisher

The Psychological Corp.

Title	<u>Columbia Mental Maturity Scale (CMMS)</u>	Age	<u>3-6 years to 9-11 years</u>
Author	<u>B.B. Burgemeister, L.H. Blum, & I. Lorge</u>	Time	<u>Untimed</u>
Copyright Date	<u>1977</u>	Type	<u>NRT</u>

Intended Purpose of Test

Intended as a quick screening device, the test measures general reasoning ability by requiring a student to make visual-perceptual discriminations in classifying and relating series of pictures, colors, forms, and symbols.

Description

An individually administered test that requires no verbal responses, the CMMS contains 92 pictorial and figural cards. The child is required to point to one drawing on each card that is different from the others. Tasks include simple perceptual classifications, such as gross differences in color, size, or form to very subtle, abstract, symbolic differences. It usually take 15 to 20 minutes to administer. This edition of the test is significantly different from earlier editions (Buros, 1978).

Norm Data

The sample consisted of 2,600 children, 100 boys and 100 girls at each 6-month interval between the ages of 3-6 and 9-11. The sample was stratified according to the U.S. 1960 census by parental occupational level, race, geographic region, age, and sex.

Reliability

Both internal consistency (split/half) and test-retest reliability were reported. Internal consistency coefficients ranged from .85 to .91. Test-retest reliability for three different age groups ranged from .84 to .86.

Validity

Correlations with other tests of intelligence (Otis-Lennon and Stanford-Binet) for students in grades 1 through 3 were in the .60's. Correlations with the Stanford Achievement Test ranged from .31 to .61.

Special Alerts/Comments

The UMMS provides a quick screening of general reasoning ability. It is important to remember that the score of the test is limited in its assessment of intellectual behavior to discrimination and classification. Therefore, inference about a student's general intellectual functioning should not be made from the test results.

Since it requires no verbal response and a minimal motor response from a student, it is especially useful with student who are cerebral palsied or speech or hearing impaired.

Spanish directions are available for administering the test. Because of the nature of the items, the author claimed that the test reflects minimal cultural biases. However, limited sampling necessitates the cautious interpretation of test scores for linguistically and culturally diverse students and LEP students. Scores should not be reported for bilingual and/or LEP students.

References

Manual
Buros, 1978
Salvia & Ysseldyke, 1985

Publisher

The Psychological Corp.

Title	<u>Culture Fair Intelligence Test (CFIT)</u>	Age	<u>4-14 years</u>
Author	<u>R. Cattell & A.K. Cattell</u>	Time	<u>12 to 22 minutes</u>
Copyright Date	<u>1960, 1973</u>	Type	<u>NRT</u>

Intended Purpose of Test

The CFIT is intended to measure "individual intelligence in a manner designed to reduce, as much as possible, the influence of verbal fluency, culture climate, and education level" (Cattell, 1972, p. 5).

Description

Designed as a group intelligence test, CFIT (Form A and B) includes three scales: Scale 1 (8 subtests) includes symbol copying, classification of pictures, mazes, identification of similar drawings (which are purported to be culture free) and selecting familiar objects when named, following directions, identifying errors in pictures of familiar objects and riddles; Scales 2 and 3 have four subtests including choosing correct pictures in a series, identifying differences, finding missing elements in a matrix, and finding relationships in pictures (Conditions: Typology).

Norm Data

Scale 1 was standardized on a combined sample of more than 400 American and British individuals. Scale 2 was standardized on 4,328 boys and girls from varied regions of the U.S. and Great Britain. The sample was apparently not stratified on the basis of any population characteristics. Scale 3 was normed on 3,140 cases of American high school students equally divided among freshmen, sophomores, juniors, and seniors and young adults in a stratified job sample.

Reliability

For Scale 1, test-retest reliability of .80 was reported for the total test. Reliability of subtests ranged from .57 to .71. For Scale 2, internal consistency ranged from .77 to .81 for Form A and .71 to .76 for Form B. Split-half reliability, ranging from .95 to .97, was computed based on a sample of 200 Mexican and American subjects. Equivalent form reliability ranged from .58 to .72 with individuals of various ages. Test-retest reliability was .82 for 200 American students and .85 for 450 11-year-old British secondary school students. No reliability data are available for the use of Scale 2 with students under 11 years of age. For Scale 3, internal consistency coefficients based on 202 high school students ranged from .51 to .68 for Form A and from .53 to .64 for Form B. Equivalent form reliability ranged from .32 to .68.

Validity

Factor-analytic studies were conducted by Cattell who claimed "the real basis of validity of an intelligence test is its correlation with the 'construct' or concept of intelligence in the general ability factor" (Salvia & Ysseldyke, 1985, p. 193). He reported that correlations of the subtests with g ranged from .53 to .99. Also reported were the correlations with the S-B ($r=.62$); the Goodenough-Harris ($r=.46$); and the WISC Full Scale ($r=.72$).

Special Alerts/Comments

As a measure of individual intellectual/cognitive ability, the results of the CFIT may be useful for screening decisions and obtaining descriptive information. Given the limitations of the standardization and reliability information, it is insufficient for determining the level of intellectual/cognitive functioning. Although a group administered test, it requires considerable expertise for proper administration and interpretation.

Because of its limited standardization, results with linguistically and culturally diverse students and LEP students should be interpreted cautiously. Scores may not be reported for bilingual and/or LEP students.

References

Manual
(Cattell-Technical Supplement,
1973)
Los Angeles Guide, 1978
Salvia & Ysseldyke, 1985

Publisher

Institute of Personality and Ability
Testing

Title Detroit Test for Learning Aptitude Age 6 - 18 years
(DTLA-2)
Author D.D. Hammill Time 50 - 120 minutes
Copyright Date 1985 Type NRT

Intended Purpose of Test

The DTLA-2 is designed to assess the general and specific learning aptitude, strengths and weaknesses of students.

Description

The DTLA-2 is a revision of the earlier edition of the DTLA. The DTLA-2 is individually administered and consists of 11 different subtests which tap motor, vocabulary, sequencing, detail recognition, and memory skills.

Norm Data

The DTLA-2 was standardized on 1,532 students from 30 states during 1984. The sample was stratified by sex, residence, race, geographic area, education of parents, ethnicity, and age. Norms tables indicate standard scores and percentiles for each subtest at 6-month intervals ranging from 6 years to 17-11 years. The number of students at each age level was not included. A norms table of composites is also included.

Reliability

Test-retest reliability estimates based on testing 33 students retested at an interval of two weeks ranged from .63 to .91 for subtests and .80 to .93 for composites. Internal consistency reliability estimates based on 150 students in the standardization sample were higher, ranging from .81 to .95 for subtests and .95 to .97 for composites.

Validity

Criterion-related validity was based on a sample of 76 students given the WISC-R and 25 given the PPVT. Results indicated a subtest median correlation of .55 and a higher composite median correlation of .71. Construct validity indicated sufficient relationship to age, group differentiation, and correlation to the SRA Achievement Series. Content validity indicated consistency with popular models of measuring intellect.

Special Alerts/Comments

The DTLA-2 provides information about learning aptitude which may be used to obtain supplementary, descriptive information with respect to learning aptitude. Given the innovative methodology for the derivation of composite scores or quotients, the results of the DTLA-2 may not be useful for screening, instructional planning, and/or evaluation of student progress decisions. Radencich (1986) added that due to lack of technical information, there is limited usefulness of its results.

The OTLA-2 is not intended to identify children whose intellectual functioning is above average" (Racencich, p. 178).

Given the low reliability of subtests, only composite scores and total score may be used for interpretation and reporting.

The author strongly stated that all the subtests should be administered in their standardized order, even if some tests are omitted.

The author indicated that due to its strong verbal-orientation, results with "poor-English speaking students" (p. 87) should be interpreted cautiously. As a result, the test is not recommended for use with LEP students and linguistically and culturally diverse students.

References

Manual
Racencich, 1986

Publisher

Pro-Ed

Title	<u>Goodenough-Harris Drawing Test (G-H)</u>	Age	<u>3 - 15 years</u>
Author	<u>D. Harris</u>	Time	<u>Untimed</u>
Copyright Date	<u>1963</u>	Type	<u>NRT</u>

Intended Purpose of Test

The G-H is designed to measure intellectual maturity, i.e., the ability to perceive, abstract, and generalize.

Description

This instrument is a brief non-verbal test administered either individually or in a group. The student is instructed to make a picture of a man. The same instructions are then given for the drawing of a woman and the drawing of the self. Two IQ's can be obtained: (1) a point scale based on credit for specific details included in the drawings and (2) a quality scale based upon a comparison of the drawing to model drawings.

Norm Data

As a revision of the Goodenough Draw-a-Man Test, the standardization sample consisted of "several thousand" students, ages 5 to 15, from four geographical areas based on the 1960 census. Equal numbers of boys and girls were included at each age level.

Reliability

Interrater reliability ranged from .80 to .96. Test-retest reliability, over a 3-month interval, was reported to range from .60's to .70's.

Validity

Concurrent validity with various intelligence tests ranged from .50 to .80. Scott (1981) reported correlations in the low .50's with the WISC-R and Stanford-Binet. Discrepancies were especially large in the average and upper ranges of intelligence. Similar correlations were found with various perceptual motor tasks (median correlations, .46).

Special Alerts/Comments

The G-H provides information about a student's ability to perceive, abstract, and generalize. Given the limited behavior sampled and the technical inadequacy of the standardization sample, the results of the G-H are insufficient for determining the level of intellectual/cognitive functioning.

It may be used to obtain descriptive information regarding a student's ability to perceive, abstract, and generalize. Since the norms and scoring are based on 1950 standards (Salvia & Ysseloyke, 1985), its appropriateness to the current New York City school population is highly questionable.

The use of the G-H with culturally and linguistically diverse students has not been demonstrated.

References

Buros, 1978
Salvia & Ysseloyke, 1985
Sattler, 1982
Scott, 1981

Publisher

The Psychological Corp.

Title	<u>Kaufman Assessment Battery for Children (K-ABC)</u>	Age	<u>2 1/2 - 12 1/2 years</u>
Author	<u>A.S. Kaufman & N.L. Kaufman</u>	Time	<u>45 - 75 minutes</u>
Copyright Date	<u>1983</u>	Type	<u>NRT</u>

Intended Purpose of Test

The K-ABC is an individually administered measure of intelligence and achievement for young and elementary school age children. It is "intended for psychological and clinical assessment, psychoeducational evaluation of learning disabled and other exceptional children, educational planning and placement, minority group assessment, preschool assessment, neuropsychological assessment, and research" (Administration and Scoring Manual, p.1).

Description

The battery consists of 16 subtests, administered according to age. No child is given more than 13 subtests. The subtests are designed to yield two types of scales, a Mental Processing Scale and an Achievement Scale. The Mental Processing Scale measures the child's ability to solve problems sequentially or simultaneously, with emphasis on the process, not on the specific content of the items. In contrast, the Achievement Scale focuses on acquired facts and applied skills. A special Nonverbal Scale, made up of those subtests that may be administered in pantomime and responded to gesturally, is also provided to assess the mental processing of hearing-impaired, speech and language disordered, and non-English speaking children.

Norms

The K-ABC was standardized on 2,000 children, 100 at each half-year of age between 2-6 and 12-5. The sample was stratified on the variables of sex, parental education, race or ethnic group, geographic region, community size, and educational placement based on 1980 U.S. Census Bureau data. Standard scores, with a mean of 100 and a standard deviation of 15, are reported for sequential, simultaneous, mental processing composite (MPC), and achievement. National percentiles, stanines, and sociocultural percentiles based upon ethnicity and parent education are also available. The battery has been adequately standardized.

Reliability

Average split-half reliability coefficients for the four Global Scales ranged from .86 to .93 for preschool children and from .89 to .97 for school-age children. Test-retest reliability coefficients, based on small samples, were generally low. The composite scales are generally reliable; the subtests are not.

Validity

The results of factor analysis offer support for the construct validity of the instrument, most notably for the existence of the two mental processes, simultaneous and sequential, hypothesized to underlie the intelligence or mental processing composite. Other validity studies reported correlations in the .80's between the Achievement Scale and the WISC-R and S-B. Correlations with mental processing subtests are lower.

Predictive validity studies reported correlations with a variety of standardized achievement tests that ranged from .02 to .89; these were highest with the Achievement Scale. Page (Mitchell, Jr., 1985 p. 773) indicated that crystallized abilities are not measured in the Mental Processing subtests and that larger loadings are found in the Achievement Tests which leads to the conclusion "that the K-ABC is mislabeling 'intelligence'" The battery does measure different mental processes, yet the validity of the battery for its intended purposes has not been clearly established.

Special Alerts/Comments

The K-ABC provides a measure of intelligence and achievement which may be useful in determining the intellectual/cognitive functioning of young children and those of elementary school age. It provides insufficient information about children who score at the upper extremes. It should only be administered, scored, and interpreted by a school psychologist.

It would seem that while empirical evidence confirms the presence of successive and simultaneous processing, its contribution to psychoeducational evaluation and instructional practices is yet to be determined. Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

References

Manuals
Naglieri, Kamphaus, & Kaufman, 1983
Salvia & Ysseldyke, 1985
Mitchell, Jr., 1985

Publisher

American Guidance Services, Inc.

Title	<u>Matrix Analogies Test-Expanded Form (MAT-EF)</u>	Age	<u>5 - 17 years</u>
Author	<u>J. A. Naglieri</u>	Time	<u>Untimed</u>
Copyright Date	<u>1985</u>	Type	<u>NPT</u>

Intended Purpose of Test

The MAT-EF is designed to measure nonverbal reasoning ability. It is intended to assess the intellectual skills of students who are nonliterate or possess limited language skills. The author emphasized the use of the MAT-EF with students who have limited English language skills, or those with cerebral palsy, deafness, or a communication disorder.

Description

The MAT-EF is administered individually. It consists of 64 abstract designs of the standard progressive matrix type, printed one per page and divided into four item groups. A screening form of the test, the MAT-Short Form (SF), is available. The MAT-SF, consisting of 34 items, provides a method of screening large numbers of students in a limited time period.

Norm Data

The MAT-EF was standardized in 1984 in two phases. In the first phase, a sample of 4,468 students was assessed on 34 of the 64 items. Class-size groups were used for this sample. The phase two procedure consisted of the individual administration of all 64 items to a sample of 1,250 students. In both phases, the sample distribution was based on 1980 U.S. Census data. The sample was stratified on the basis of geographic, sex, and ethnic group representation. Students ranged in age from 5 to 17.

Reliability

Internal consistency coefficients using Cronbach's alpha ranged from .88 to .95 for the total test with a median coefficient of .93.

Validity

Construct validity was determined from an investigation of the age-to-age progression of mean scores through factor analysis. Since solving the item of the MAT-EF is expected to increase with age, increases in mean test scores with age provide evidence of construct validity.

Criterion-related validity was established by examining the correlation between the MAT-EF and the WISC-R Performance Scale and Raven's Progressive Matrices.

Special Alerts/Comments

The MAT-EF, designed to measure non-verbal reasoning ability, may be useful in obtaining supplementary, descriptive information about a student's cognitive abilities. However, it is insufficient for determining the level of intellectual/cognitive functioning.

The MAT-EF may be valuable in assessing the cognitive development of linguistically and culturally diverse students and LEP students. However, cautious interpretation of these students' results is recommended.

Reference

Manual
Raven, 1956

Publisher

The Psychological Corp.

Title	<u>McCarthy Scales of Children's Abilities (MSCA)</u>	Age	<u>2 1/2 to 8 1/2 years</u>
Author	<u>D. McCarthy</u>	Time	<u>50 minutes to 1 hour</u>
Copyright Date	<u>1970, 1972</u>	Type	<u>NRT</u>

Intended Purpose of Test

The MSCA are designed to assess young children's general intellectual level as well as their strengths and weaknesses in several ability areas.

Description

The test kit contains 18 tests grouped into one or more of six scales. Five verbal tests, seven perceptual-performance tests, and three quantitative tests are included in the General Cognitive Scale. There are also tests of gross motor coordination. The test is individually administered.

Norm Data

The standardization sample consisted of approximately 1,000 children, 100 at each of ten age groups, 2 1/2, 3, 3 1/2, 4, 4 1/2, 5, 5 1/2, 6 1/2, 7 1/2, and 8 1/2. The sample was equally divided for sex and was selected from four major geographic regions: Northeast, North Central, South, and West. The sample excluded students with marked handicaps, bilingual students, and LEP students.

Reliability

Reliability coefficients and standard errors of measurement were reported for ten different age levels for each of the six scales. The average split-half reliability of the General Cognitive Index was .93, while for the other five scales coefficients ranged from .79 to .88. Test-retest reliability based on a 30-day interval ranged from .69 to .89 for the various subtests.

Validity

Concurrent validity studies using the S-B, WISC, WISC-R, and WPPSI were reported. Correlations of the MSCA General Cognitive Scale and IQ's obtained from the other instruments ranged from .62 to .81. Lowest correlations were reported for the Motor Scale ranging from .45 to .91 (median .85). Predictive validity, using the Metropolitan Achievement Test (N=31), ranged from .03 to .54.

Special Alerts/Comments

The MSCA provide measures of intellectual/cognitive levels which may be useful in determining the intellectual/cognitive functioning of young children.

The scoring procedures are concise and straightforward. It is easy to administer and the tasks are enjoyable.

The tasks on the MSCA do not provide sufficient information for discriminating severe mental retardation and giftedness. The usefulness of the test with severely handicapped students, bilingual students, and LEP students is unsubstantiated and caution should be exercised when determining the intellectual/cognitive level of such students.

Scores should not be reported for bilingual and/or LEP students.

References

Buros, 1978
Salvia & Ysseldyke, 1985
Sattler, 1982

Publisher

The Psychological Corp.

Title	<u>Nebraska Test of Learning Aptitude (NTLA)</u>	Age	<u>3 - 16 years</u>
Author	<u>S. Hiskey</u>	Time	<u>50 - 60 minutes</u>
Copyright Date	<u>1966</u>	Type	<u>NRT</u>

Intended Purpose of Test

The NTLA, formerly known as the Hiskey-Nebraska, is designed to assess learning aptitude of deaf and hearing individuals.

Description

The NTLA is an individually administered test consisting of 12 subtests. Pantomime administration of the test is available for deaf students and verbal directions for use with hearing students. Each of the 12 subtests is a power test: bead patterns, memory for color, picture identification, picture association, paper folding, visual attention span, block patterns, completion of drawings, memory for digits, puzzle blocks, picture analogies, spatial reasoning. The student earns points on each of the subtests administered. Different subtests employ different ceiling rules and criteria which are described in the manual.

Norm Data

Norms for hearing students were first published in 1957 and the revised edition has norms for both deaf and hearing students. In the 1966 restandardization, there were 1,079 deaf students and 1,074 hearing students from ten states extending from New York to Utah to Florida. Samples included representatives from minority groups but specific percentages were not reported. No breakdown by race or sex was included in the manual. The deaf students were from state schools for the deaf, but the number of students from each school was not given. Hearing students were selected by occupational levels of parents according to the 1960 census. The 3- and 4-year-old samples of deaf students and samples of older hearing students were limited in size.

Reliability

Split-half reliabilities of .95 for 3- to 10-year-old deaf students, .92 for 11- to 17-year-old deaf students, .93 for 3- to 10 year-old hearing students, and .90 for 11- to 17-year-old hearing students were reported in the manual. No data on the standard error of measurement were included. There were no reported data for the reliability of individual subtests.

Validity

Concurrent validity coefficients for the 1966 revision were reported as .86 for 99 hearing students (ages 3 to 10) and .78 for 50 hearing students (ages 11 to 17) with the Stanford-Binet and .82 with the WISC for 52 hearing students (ages 5 to 11).

Special Alerts/Comments

The NTLA provides information about the learning aptitude of deaf and hearing individuals which may be useful for screening decisions with hearing impaired students between the ages of 5 and 12. Given its limited technical data, the results are insufficient for determining the level of intellectual/cognitive functioning without support from other measures of intellectual/cognitive functioning.

The examiner must have considerable experience in individual intellectual assessment and also, if possible, experience working with the deaf. The manual contains a discussion of desirable practices to be followed in testing deaf children.

Given the limitations of the standardization sample, use of the NTLA with culturally and linguistically diverse students and LEP students has not been established. Scores should not be reported for bilingual and/or LEP students.

References

Manual
Anastasi, 1976
Buros, 1978
Salvia & Ysselcyke, 1985

Publisher

Marshall S. Hiskey

Title	<u>Progressive Matrices (Raven)</u>	Age	<u>6 years to adult</u>
Author	<u>J.C. Raven</u>	Time	<u>Untimed</u>
Copyright Date	<u>1938, 1947, 1956</u>	Type	<u>NRT</u>

Intended Purpose of Test

The test is designed to provide a culture-free measure of intellectual functioning.

Description

This test is individually or group administered and comes in three different forms according to age. The Coloured Matrices Form is designed especially for children. "In each form, the child is presented with a matrix-like arrangement of figural symbols, and must complete the matrix by selecting the appropriate missing symbol from a group of symbols. Raw scores are converted to percentiles" (Manual, p. 7). Very little oral instructions are required.

Norm Data

The Raven was standardized on a British sample of approximately 7,200 individuals ranging in age from 6 to 65. Percentile norms are provided in the manual for each half-year interval between 8 and 14 years.

Reliability

Test-retest reliabilities reported in recent years (Anastasi, 1976) ranged between .70 to .90.

Validity

The Raven has been supported in factor-analytic studies as a measure of Spearman's g ; however, other factors have been found which also influence the test score. Correlations, reported in several studies, with the Stanford-Binet and the WISC ranged from .55 to .86.

Special Alerts/Comments

The Raven provides a measure of intellectual/cognitive functioning which may be useful in obtaining supplementary, descriptive information about a student's cognitive abilities. However, it is insufficient for determining the level of intellectual/cognitive functioning.

It should not be used with students exhibiting visual perceptual difficulties. Administration of the Raven should be conducted by a psychologist experienced in intelligence testing.

As a non-verbal test, it may be appropriate for a broad range of populations including students from culturally and linguistically diverse backgrounds as well as LEP students.

Because of the limited standardization, scores should not be reported for any students. Results should be presented descriptively.

References

Manual
Anastasi, 1976
Los Angeles County, 1978
Sattler, 1982
Tarczan, 1972

Publisher

The Psychological Corp.

Title	<u>Stanford-Binet Intelligence Scale (S-B)</u>	Age	<u>2 - Adult</u>
Author	<u>R.L. Thorndike, E.P. Hagen, & J.M. Sattler</u>	Time	<u>60-90 minutes</u>
Copyright Date	<u>1986 (4th Edition)</u>	Type	<u>NRT</u>

Intended Purpose of Test

The S-B is designed to assess the overall level of cognitive development as well as the pattern of this development. It is constructed to help identify gifted students, as well as to help differentiate between students who are mentally retarded and those who have specific learning difficulties.

Description

The newest edition of the S-B is a major revision of the 1960 test. It is individually administered. The new Stanford is similar to the older edition in that it requires establishing basal and ceiling levels. Many of the same items are included. However, the age-scale format has been replaced by tests, each using a single item type.

The revision consists of 15 tests, tapping different cognitive skills. Appropriate age levels for each test have been determined. The four broad categories of the new S-B are: Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning, and Short-Term Memory. The evaluation of general reasoning ability, or g, is retained. In addition, scores can now be generated for any of the four areas described above, any combination of these four categories, and individual subtest scores.

The advantages of the new S-B are: broader coverage of cognitive skills and information processing abilities, more flexibility in test administration, higher reliability and test-score precision, and a more detailed diagnostic assessment of cognitive and information processing skills.

A complete S-B consists of 8 to 13 subtests, depending upon the student's chronological age.

Norm Data

Norms were developed from 5,000 subjects based on 1980 U.S. Census data. The design for standardization was based on geographic region, community size, ethnic group, age, and gender. Standardization took place in 1985. The norm tables included in the Manual provide for conversions of raw scores to Standard Age Scores (SAS) for subtests, areas, and composite.

Reliability

Kuder-Richardson Formula 20 procedures were used to determine internal consistency of the S-B. Median reliabilities for the fifteen tests ranged from .73 to .92. The most reliable score is the Composite Standard Age

Score. It is therefore recommended that this score be the primary source of information for making decisions. The next most reliable scores are the Area SASs.

Test-retest reliability was obtained by retesting 112 students over a 2- to 8-month interval. For the Composite Score, test-retest reliability estimates were reported as .90 for 8-year-olds and .91 for 5-year-olds.

Validity

Construct validity was determined by investigating the correlations between the S-B and WPPSI, WISC-R, WAIS-R, and the K-ABC; the obtained correlation coefficients provided strong support for the construct validity of the S-B.

Special Alerts/Comments

The S-B provides an overall indication of intellectual abilities. It may be useful in determining the intellectual/cognitive functioning of school age students. The Fourth Edition may be useful for differentiating between students who have specific learning disabilities and those who are mentally retarded through a comparison of individual or combinations of cognitive ability areas. However, scatter between subtests and/or areas must be significant; i.e., standard score differences must be greater than one standard deviation.

An abbreviated test battery is not acceptable for determining a level of intellectual/cognitive functioning in an initial referral; in order to be used for this purpose, a standard battery of 8 to 13 subtests must be administered to an individual student.

The S-B should only be administered, scored, and interpreted by a school psychologist.

It has been recommended that a subtest should never be dropped from estimating a total scores unless the score is zero.

The S-B is not recommended for use with linguistically and culturally diverse students and LEP students. Scores should not be reported for bilingual and/or LEP students.

References

Manual
Technical Manual

Publisher

Riverside Publishing Co.

Title	<u>System of Multicultural Pluralistic Assessment (SOMPA)</u>	Age	<u>5 to 11</u>
Author	<u>R. Mercer & F. Lewis</u>	Time	<u>3 hours (approx.)</u>
Copyright Date	<u>1977</u>	Type	<u>NRT</u>

Intended Purpose of Test

The SOMPA is a system to provide nondiscriminatory assessment of cognitive and perceptual motor skills, adaptive behavior, and general physical well-being, and to develop an estimated learning potential (ELP).

Description

The SOMPA is an assessment system using three different models simultaneously; the models are the medical model, the social system model, and the pluralistic model. Within each assessment model there are a variety of measures. The medical model includes physical dexterity tasks, the Bender Gestalt scored according to the Koppitz system, weight by height, visual acuity, auditory acuity, and health history inventories. Measures within the social system model are the WISC-R and the ABIC. In the pluralistic model, the measures are Sociocultural Scales, consisting of family size, family structure, socioeconomic status, and urban acculturation, which are used to adjust WISC-R scores to obtain ELP. The ELP indicated a student's WISC-R scores relative to his/her sociocultural and ethnic peers.

Norm Data

The SOMPA measures were developed and standardized on 2,085 California public school students, 5 through 11 years of age during the 1972-1973 school year. Representative sub samples include 697 black, 699 English-speaking and 690 Spanish-surnamed students. Approximately half were male and half female.

Reliability

Each measure in the SOMPA has a separate reliability estimate. Reviews for the WISC-R, ABIC, and Bender Gestalt (See other sections of the Guide) include the reliability of these tasks. Split-half reliability coefficients for physical dexterity tasks at each age level ranged from .60 to .94. No reliability data were presented for the Sociocultural Scales.

Validity

Reviews for the WISC-R, ABIC, and Bender Gestalt (See other sections of the Guide) include validity information for these instruments. Construct validity of the Sociocultural Scales was demonstrated through factor analysis which yielded nine significant factors, grouped into four scales.

These provide a substantial contribution to the variance of the verbal IQ (19.2%), performance IQ (6.5%), and full scale IQ (14.4%). Mean scores of the nine factors and four scales differed significantly between ethnic groups. Content validity was developed for the physical dexterity tasks.

Special Alerts/Comments

The SOMPA provides a system to assess cognitive functioning, perceptual motor skills, adaptive behavior, and general well-being which may be used to obtain supplementary, descriptive information. Given the limited standardization of the system, the results are insufficient for determining the level of intellectual/cognitive functioning.

The WISC-R and Bender Gestalt must be administered, scored, and interpreted by a school psychologist. The physical dexterity tasks, weight by height, and visual acuity require administration by a specially trained nurse. Auditory acuity should also be administered and scored by a nurse or audiometrist.

The SOMPA procedure is based on the assumptions that: (1) the WISC-R is an achievement test, (2) the difference between groups is an artifact of the cultural specificity of the test and the exposure to Anglo core culture, and (3) the Sociocultural Scales indicate the student's location in sociocultural space which may be used to adjust WISC-R scores. However, this adjustment is inappropriate for students 5 to 6 1/2 years of age since the WISC-R is not recommended for students below 6 1/2 years. It has been strongly recommended that since the SOMPA procedure was based on a California population, local norms should be developed and the sensitivity of the Sociocultural Scales be determined for the New York City population. Based upon available literature (Goodman, 1979; Reschly, 1980), the procedure should be used cautiously with students in New York City.

Scores should not be reported for bilingual and/or LEP students.

References

Publisher

Manual
Figueroa, 1979
Goodman, 1979
Reschly, 1980

The Psychological Corp.

Title	<u>Vane Kindergarten Test (VKT)</u>	Age	<u>4 1/2 to 6 years</u>
Author	<u>J.R. Vane</u>	Time	<u>Untimed (approx. 20-30 mins.)</u>
Copyright Date	<u>1968</u>	Type	<u>NRT</u>

Intended Purpose of Test

The VKT provides measures of intelligence and behavior adjustment for young students.

Description

The test contains three subtests: Perceptual Motor, Vocabulary, and Man (draw-a-person). The Perceptual Motor and Man subtests may be administered to groups of 10 to 12 students. The Vocabulary subtest is administered individually by a school psychologist. Scoring keys are provided.

Norm Data

The standardization sample consisted of 800 students, equally divided by sex, ranging in age from 4 1/2 to 6, and drawn from the states of New York and New Jersey. The ethnic and socioeconomic distributions were similar to that of the total U.S. population and included urban and rural representation.

Reliability

Two studies of test-retest reliability were conducted with 14 and 36 kindergarten students; coefficients of .97 and .88, respectively, were reported. Subtest reliabilities ranged between .71 and .92.

Validity

Two separate studies reported correlations of .76 with an earlier edition of the S-8. Coefficients of about .60 were reported with first, second, and third grade achievement tests.

Special/Alerts Comments

The VKT provides information about a young student's perceptual motor, vocabulary, and draw-a-person abilities useful for obtaining supplementary, descriptive information about such abilities. Given its technical inadequacies and limited samples of intellectual behavior, the results of the VKT are insufficient for determining intellectual/cognitive ability.

The test should be administered by a school psychologist. Given the restrictive nature of the sample, its use with LEP students and linguistically and culturally diverse students is highly questionable.

References

Manual
Buros

Publisher

Clinical Psychology Publishing Comp.

Title	<u>Wechsler Adult Intelligence Scale</u> <u>Revised (WAIS-R)</u>	Age	<u>16 years +</u>
Author	<u>D. Wechsler</u>	Time	<u>Average of 1 hour</u>
Copyright Date	<u>1981</u>	Type	<u>NRT</u>

Intended Purpose of Test

The WAIS-R is designed to measure general intelligence in adults:

Description

WAIS-R is a revision of the 1955 WAIS, allowing replacement of outdated items. Eighty percent of the original items were retained. It is an individually administered test consisting of two scales which contain five Performance subtests and six Verbal subtests. Three IQ scores may be obtained: Verbal, Performance, and Full Range. The manual provides detailed directions for administration and scoring.

Norm Data

The standardization sample, 1,880 subjects, was based on the distribution of the population of the U.S. according to the 1950 census. The sample was stratified on the following variables: age, sex, race, geographic region, occupation, education, and urban-rural residence.

Reliability

Split-half reliabilities averaged .97 for the Full Scale, .96 for the Verbal Scale and .93 for the Performance Scale for all age groups. Reliability coefficients for the subtest ranged from .60 to .96. Test-retest coefficients were presented by age group and averaged .97 for the Full Scale IQ. Standard errors of measurement (SEM) were 2.74, 4.14, and 2.53 for Verbal, Performance, and Full Scale, respectively.

Validity

Since the WAIS-R measures the same abilities as the 1955 WAIS and the 1938 Wechsler-Bellevue Scale, evidence of the validity of these prior scales was considered relevant to the WAIS-R. Correlations with performance in school were about .50. Factor analysis studies have identified three factors termed verbal comprehension, perceptual organization, and freedom from distractibility (sometimes termed memory) corresponding generally to the scale divisions on the WAIS.

Special Alerts/Comments

The WAIS-R provides a measure of general intelligence of adults. The test findings are useful for determining intellectual/cognitive ability of older students. Administration, scoring, and interpretation are restricted to school psychologists.

Numerous studies are available describing the clinical and educational uses of the WAIS-R. A Spanish version of the WAIS is reviewed in the LEP assessment section of this Guide.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously. Scores may not be reported for bilingual and/or LEP students.

References

Manual
Zimmerman & Woo-Sam, 1973
Salvia & Ysseldyke, 1985

Publisher

The Psychological Corp.

Title	<u>Wechsler Intelligence Scale for</u> <u>Children Revised (WISC-R)</u>	Age	<u>6 - 16 years</u>
Author	<u>D. Wechsler</u>	Time	<u>Average of 1 hour</u>
Copyright Date	<u>1974</u>	Type	<u>NRT</u>

Intended Purpose of Test

The WISC-R is designed as a measure of global intelligence defined as "the overall capacity of an individual to understand and cope with the world around him. It conceives of intelligence as an overall or global entity; that is, a multidetermined and multifaceted entity... [Intelligence] "is inferred from the way these abilities are manifested under different conditions and circumstances" (Wechsler, 1974).

Description

The test consists of a Verbal and Performance Scale, each containing six subtests (one optional subtest in each scale). The test is individually administered and has scaled scores for each subtest; IQ scores for each scale and a Full Scale IQ may be computed.

Norm Data

The test was standardized on 2,200 white and non-white American students. Blacks, American Indians, Orientals, Puerto Ricans, and Mexican-Americans in approximate proportion to the population on the 1970 census were included. Students were also distributed evenly over the age groups from 6-16 years. Other variables included proportional distributions on sex, geographic region, parent occupation, and urban-rural residence.

Reliability

The three scales of the test (Verbal, Performance, Full) had reliability coefficients of .89 or above over the entire age range. Subtest reliabilities ranged from .62 to .92. The reliability of the total test score has been considered outstanding.

Validity

The WISC-R correlated .82 (Full Scale) with an earlier edition of the S-B. Salvia and Ysseloyke (1985) reported somewhat lower correlations with the S-B among small groups of students at four age levels. Correlations with overlapping portions of the WAIS and WPPSI were .95 and .82, respectively. Correlations with achievement tests (PIAT, WRAT) averaged .60 and with school grades, .39.

Special Alerts/Comments

The WISC-R provides a measure of global intelligence of students. It may be useful in determining the intellectual/cognitive functioning of school age students. Administration, scoring, and interpretation are restricted to school psychologists.

Numerous studies regarding the clinical and educational use of the WISC-R are available. In New York City, the findings of the Project REP indicated that the full scales should be administered in determining intellectual/cognitive ability.

Standardization of the WISC-R Performance Scale (Anderson & Sisco, 1977) for 1,228 deaf students resulted in statistically significant but minor differences between scores of hearing and deaf students; therefore, WISC-R Performance norms for hearing students may be considered adequate for the deaf population.

A Spanish edition of the WISC is reviewed in the LEP assessment section of the Guide. The verbal scale is insufficient for determining the intellectual/cognitive level of LEP students.

Norms are not available for bilingual and LEP students, therefore, scores may not be reported when the test is administered to them in English or Spanish.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

References

Manual
Anderson & Sisco, 1977
Glasser & Zimmerman, 1967
Kaufman, 1979
Salvia & Yssedyke, 1985
Wechsler, 1974

Publisher

The Psychological Corp.

Title	<u>Wechsler Preschool and Primary Scale of Intelligence (WPPSI)</u>	Age	<u>4 to 6-6 years</u>
Author	<u>D. Wechsler</u>	Time	<u>50 - 75 minutes</u>
Copyright Date	<u>1963, 1967</u>	Type	<u>NRT</u>

Intended Purpose of Test

The WPPSI is designed to measure general intelligence of young children.

Description

Similar to the WISC-R and the WAIS-R, the WPPSI contains five Verbal tests plus an alternate and five Performance tests. It yields a Verbal, Performance, and Full Scale score. It is individually administered.

Norm Data

The sample consisted of 1,200 children, 100 boys and 100 girls in each of six age groups by half years. The sample was stratified based on the 1960 census taking into account race, sex, urban-rural residence, and father's occupation.

Reliability

Split-half reliabilities were reported for every subtest at each age level and ranged from .93 to .96. In addition, test-retest reliabilities for all but one subtest were in the .80's. These data were higher than those reported by Salvia & Ysseldyke (1985).

Validity

Concurrent validity was indicated through correlations between the WPPSI Full Scale and the S-B and the PPVT, which were .75 and .58, respectively.

Special Alerts/Comments

The WPPSI provides a measure of intellectual/cognitive ability of very young children which may be useful in determining their intellectual/cognitive functioning. It provides little information about students who score at either lower or upper extremes. Like the WAIS-R and WISC-R, the WPPSI may only be administered, scored, and interpreted by a school psychologist. Although young children find the material interesting, testing often requires two sessions (Buros, 1972).

It is not recommended for use with linguistically and culturally diverse students and LEP students. Scores should not be reported for bilingual and/or LEP students.

References

Buros, 1972
Salvia & Ysseldyke, 1985
Sattler, 1982

Publisher

The Psychological Corp.

Title	<u>Woodcock-Johnson Psychoeducational Battery (W-J)</u>	Age	<u>3-80 years</u>
Author	<u>R.W. Woodcock & M.B. Johnson</u>	Time	<u>2-3 hours</u>
Copyright Date	<u>1978</u>	Type	<u>NRT</u>

Intended Purpose of Test

The W-J Battery is designed to measure cognitive abilities, scholastic aptitudes, achievement, and interests. The authors stated that "...the major uses of the Battery include individual assessment, selection and placement, individual program planning, guidance and research" (Manual, 1978, p.4).

Description

The complete W-J Battery is packaged in notebook-like kits. It consists of a mixture of multiple-choice and open-ended items requiring both oral and written responses and is divided into three sections: cognitive ability, achievement, interest level. There are a total of 27 subtests. In addition to grade level and age equivalent scores, functional level scores can also be obtained from the subtest results. Subtest scores may be grouped into clusters. Achievement-aptitude differences are also reported.

Norm Data

Normative data were gathered in 1976 and 1977 on 4,732 subjects in 42 communities throughout the U.S. Weights were assigned to equal the sex-by-race and region-by-urbanization distribution of the U.S., as reported by the Bureau of the Census. Handicapped students were excluded from the norm sample unless they were being educated within regular education classes.

Reliability

Split-half reliabilities for individual tests ranged from .60 to .90. Over 200 test-retest reliability coefficients of subtest scores at varying grade levels were primarily between .80 and .90. These coefficients were high even among preschoolers.

Validity

The manual reported several concurrent validity studies which showed moderate correlations (.54 to .83) between the W-J cognitive objectives and the PPVT, S-B, and WISC-R. Correlations between W-J achievement and the PIAT, WRAT, and Key Math ranged from .46 to .92. The W-J Battery correlations for predicting school achievement ranged from .45 to .86. A variety of concurrent and predictive validity studies were undertaken with third, fifth, and twelfth graders. Correlations with the PIAT, Iowa, and Woodcock Reading Mastery Test for these grades ranged from .70 to .90.

for the reading battery. Correlations of the math battery with the Iowa and Key Math ranged from .60 to .80. The written Language tests correlated with the PIAT and Iowa between .70 and .80. Preschool cognitive ability clusters correlated .83 with the Stanford-Binet, .54 with the Denver, .60 with the DIAL, and, .70 with the PPVT. Correlations of third, fifth, and twelfth grade students' cognitive ability scores with the Full Scale WISC-R, Verbal WISC-R, and Performance WISC-R were .70, .75, and .58, respectively. With the WAIS the correlations were, respectively, .83, .81, and .67.

Special Alerts/Comments

The W-J battery may be useful for screening and evaluation of student progress in reading and arithmetic and written language. It also can provide descriptive information with regard to verbal cognitive ability. Bracken, Prasse, and Breen (In press) concluded that the WISC-R is preferred over the W-J Battery. This conclusion supported findings of differences between W-J Battery scores and WISC-R scores for learning disabled students (Reeve, Hall, & Zakreski, 1979; Algozzine, Ysseldyke, & Shinn, 1980).

Give these research findings and the need for information on cognitive functioning which considers nonverbal behavior as well as verbal behavior, the results of the W-J Battery are insufficient for determining a student's level of intellectual/cognitive functioning. However, the information provided may be useful in determining performance levels in reading and math and the extent of the discrepancy between achievement and aptitude.

Because of the limited standardization, its use with culturally and linguistically diverse students and LEP students has not been established. Scores should not be reported for bilingual and/or LEP students.

References

Publisher

Technical Report
Manual
Reeve, Hall, & Zakreski, 1979
Algozzine, Ysseldyke,
& Shinn, 1980
Salvia & Ysseldyke, 1985
Bracken, Prasse, & Breen, (In Press)

DLM Teaching Resources

PERCEPTUAL AND MOTOR FUNCTIONING

These tests measure skills in perception (visual, auditory), integration of perceptual abilities, and fine and gross motor functions. Some tests provide an indication of possible neurological impairment.

Caution should be taken when using these tests with physically handicapped and/or linguistically diverse students.

From a developmental perspective, they are primarily intended to provide information about the early stages of development.

Title The Assessment of Phonological
Processes (APP)

Age 2 - 9 years

Author B.W. Hoosun

Time 20 minutes

Copyright Date 1980

Type IRT

Intended Purpose of Test

The APP is designed to assess the basic phonological processes involved in speech sound production.

Description

The APP kit consists of a manual and four sets of form. The recording form provides target phonetic transcriptions for 55 words with spaces to recording deviant responses; the analysis form provides the test user with a format to identify phonological processes; a summary form provides space for results and information on occurrence percentages for various major phonological processes; a screening form provides a quick analysis of the twenty most common phonological processes.

Norm Data

Normative data are presented for a sample of 60 4-year-olds. No information is provided on subject selection. The method by which the norms were gathered must be considered as not complying with professional standards. Although the author recommended using the APP for 2- and 3-year-olds and older students, the norms were developed only on data from 4-year-olds.

Reliability

No reliability was reported.

Validity

No information was provided.

Special Alerts/Comments

The APP provides information on basic phonological processes which may be useful for descriptive and supplementary data. Based upon inadequate technical information the results are insufficient for making screening, performance level, instructional planning, and evaluation of student process decisions. It is not recommended for use with linguistically and culturally diverse students and LEP students.

Until better norms and more descriptive information in the normative processes are determined use of the test should be extremely guarded at best.

Reference

Buros, 1985

Publisher

Interstate Printers and
Publishers, Inc.

Title	<u>Clinical Evaluation of Language Functions (CELF)</u>	Grade	<u>K - 12th</u>
Author	<u>E. Semel & E. Wiig</u>	Time	<u>Timed and Untimed (approx. 90 minutes)</u>
Copyright Date	<u>1980</u>	Type	<u>NRT</u>

Intended Purpose of Test

The purpose of the CELF is to provide a measure of the language processing and production abilities of school-age students.

Description

The CELF consists of processing and production subtests yielding a total score based on the sum of the raw scores of the two subtests. The test has two levels, elementary and advanced; the elementary level contains 31 processing items and 17 production items; the advanced level, 34 and 18, respectively. Processing items were designed to screen for significant delays and deficits in aspects of language processing related to perception, recognition, recall, and interpretation of spoken language. Production items screen for agility and accuracy in phoneme production, ability to recall, identify, and retrieve words and concepts; accuracy of serial recalls; and immediate recall of model sentences.

Norm Data

The test was standardized on 634 students for the elementary level and on 777 for the advanced level. The sample was stratified to reflect the 1970 U.S. census with respect to sex, residence, geographic region, ethnic background, and parent occupation.

All scores are converted to percentile ranks which have been developed for each age level.

Reliability

Internal consistency reliability coefficients for the elementary form were reported as .89 for Processing, .87 for Production, and .82 for Total score. For the advanced form, they were .74, .62, and .78, respectively. Test-retest reliability coefficients were generally in the .70's and .80's.

Validity

Concurrent validity, based on correlations with the Illinois Test of Psycholinguistic Abilities (Verbal), DTLA (Verbal), and Northwestern Syntax Screening Test (Receptive and Expressive) was, respectively, .62, .49, and .48 for the elementary form, and .46, .55, and .47 for the advanced form.

Special Alerts/Comments

The CELF provides a gross measure of the language processing and production abilities of students. The results of the CELF may be useful for making screening decisions to determine further areas of oral language where assessment is needed. It is insufficient for making instructional planning or evaluation of student progress decisions and/or developing objectives on the IEP in the area of Communication and Language.

It is not recommended for use with LEP students.

References

Manual
Ritner, In Press

Publisher

The Psychological Corp.

Title Fisher-Logemann Test of
Articulation Competence

Age 6 years - Adult

Author H.B. Fisher & J.A. Logemann

Time Untimed

Copyright Date 1971

Type CRT

Intended Purpose of Test

The Fisher-Logemann is designed to assess systematically an individual's phonological system.

Description

The Fisher-Logemann is available in two forms. One form elicits spontaneous responses to picture stimuli. It consists of 109 pictures on 35 hard-board cards in a spiral folio; the folio may be easily converted into an easel for convenient administration. A selected portion comprises a rapid screening test. The other test form consists of 15 sentences to be articulated by the student.

The manual, which provides technical, administrative and scoring information for both the picture and sentence tests, includes references for dialects and foreign language. Response forms for both test forms provide space for noting information about dialectical variations and foreign language spoken in the home. Both test forms are arranged so that phonemic entries are categorized by place of articulation, manner of formation, and by the feature of voicing. Suggested prompt phrases are provided for the examiner if no response is elicited by the picture stimuli.

Test Administration

The Fisher-Logemann is individually administered. It is recommended that the examiner be thoroughly familiar with all phonemes, as well as the method of articulation, formation and voicing. No special materials are necessary.

Technical Data

The Fisher-Logemann was developed out of a need for isolating more specific phoneme difficulties than identified in current articulation tests.

The test was piloted on 500 students of varying dialectical backgrounds and socioeconomic levels in the Chicago area. The majority of these students were in the lower elementary grades, including Kindergarten. Children with various handicapping conditions and varied speech problems were also included.

No information is available regarding reliability or validity.

Effective Use/Comments

The Fisher-Logemann is designed to assess systematically an individual's phonological system and provide an organized framework to record and analyze phonetic notations of misarticulation. Although the instrument may be useful for obtaining descriptive information about the student's in communication and language and/or speech, the test findings are insufficient for making individual performance level, evaluation of students process and instructional planning decisions and/or developing IEP goals in these areas.

Both the picture and sentence forms may be used with children as well as adults. The picture test is especially appropriate for use with very young children or mentally retarded and distractible subjects. The sentence test has been used successfully with as early as third grade readers. A tape recording of the test is strongly recommended.

It should only be given and analyzed by professionals trained in speech pathology. Since the manual provides explicit information regarding dialects and foreign languages, it may be used with linguistically and culturally diverse students and LEP students; however, results should be interpreted cautiously.

References

Manual
Boros, 1978

Publisher

Riverside Publishing Co.

Title Fullerton Language Test for Age 11 - 18 years
Adolescents (FLTA)
Author A.R. Thorum Time Untimed (45 minutes)
Copyright Date 1980 Type NRT

Intended Purpose for Test

The FLTA was developed to distinguish between normal and language impaired adolescents. It assesses receptive and expressive oral language.

Description

The FLTA consists of eight subtests: Auditory Synthesis, Morphology Competency, Oral Commands, Convergent Production, Divergent Productions, Syllabication, Grammatic Competency, and Idioms. The elements tested are phonemes, syllables, suffixes, prefixes, syntactical structures of varying complexities, homonyms, classification systems, grammatical structures, and idioms.

The test is individually administered in as many sessions as necessary. Subtests may be given in any order and no special training or materials are required. The test is untimed except for the subtest on divergent production.

Norm Data

The standardization sample consisted of 762 students, ages 11 through 18, from regular classrooms in cities and rural areas of California and Oregon. Most students were white. Black, Spanish, and other ethnic groups were represented by small percentages.

Results are reported in terms of raw scores which are converted to Competency Level, Instruction Level, and Frustration Level.

Reliability

Test-retest reliability for individual subtests established through Pearson product-moment correlations ranged from .84 to .96. Internal consistency was established using the Kuder-Richardson method and resulted in coefficients ranging from .70 to .85.

Validity

Diagnostic validity studies indicated a significant level of difference between scores attained by a special education population and a mainstream population. High intercorrelations between subtests indicated that each measured a related component of language.

Special Alerts/Comments

The FLTA provides information on a student's receptive and oral expressive language. Due to the limitations of the standardization sample, the information provided is insufficient for making screening, performance level and instructional planning decisions. With other instruments, it may be used for developing shortterm objectives on the IEP.

The test's usefulness has not been demonstrated with linguistically and culturally diverse students and LEP students. Results with these populations should be interpreted with extreme caution. Scores should not be reported for LEP and/or bilingual students.

Reference

Manual

Publisher

Consulting Psychologists Press, Inc.

Title Goldman-Fristoe Test of Articulation
(GFTA)

Age 6-16+ years

Author R. Goldman & M. Fristoe

Time Untimed (15 minutes)

Copyright Date 1969, 1972

Type NRT

Intended Purpose of Test

This test provides a systematic means of assessing an individual's articulation of the consonant sounds.

Description

The GFTA is individually administered and consists of three subtests: sounds-in-words in which 35 pictures are presented to the student; sounds-in-sentences designed to elicit complex and spontaneous speech; and stimulability to stimulate correct production of the misarticulated sounds.

Norm Data

Percentile ranks were based on data collected from two out of three subtests, sounds-in-words and stimulability, in a national survey of 38,884 students, ages 6 to 16+.

Reliability

Using the percent-agreement method, three types of reliability were reported: test-retest, interrater, and intrarater. All correlation coefficients were above .90.

Validity

The authors stated that the collection of items used in this test to assess speech sound production assures its content validity.

Special Alerts/Comments

The GFTA provides an indication of an individual's articulation of consonant sounds. It may be useful in obtaining supplementary, descriptive information about a student's articulation of consonant sounds. The test findings are insufficient for making performance level, instructional planning, and evaluation of student progress decisions.

It should only be given and analyzed by professionals trained in speech pathology. Because of articulation and phonology differences in speech production, it is not appropriate for LEP students.

References

Manual
Buros, 1978

Publisher

American Guidance Service, Inc.

Title	<u>Grammatical Analysis of Elicited</u> <u>Language-Complex Sentence Level</u> <u>(GAEL-C)</u>	Age	<u>3-0 to 5-11 years</u>
Author	<u>J.S. Moog & A.E. Geers</u>	Time	<u>45 minutes</u>
Copyright Date	<u>1980</u>	Type	<u>NRT</u>

Intended Purpose of Test

The GAEL-C is designed to evaluate the child's acquisition of specific grammatical structures. It may be used with normal-hearing, hearing-impaired, or profoundly deaf children.

Description

The GAEL-C consists of 22 activities which elicit 88 target sentences, illustrating various complex sentence structures. The activities include toy manipulation, story telling, picture description, and a variety of games. Each of the target sentences is at least six words long and contains more than one verb form.

Detailed instructions for test administration and scoring are included in the manual. Results are reported in terms of raw scores, standard scores, and percentiles.

The complete kit includes a manual, test booklets, transcription sheets, score sheets, and summary sheets/record forms.

The test is individually administered and the child's responses are tape recorded.

Norm Data

The GAEL-C was standardized on 510 children; 240 had normal hearing; 120 were severely hearing impaired; 150 were profoundly deaf. The normal-hearing children were from 17 different pre-school and day care centers, while those with impaired hearing or profound deafness were between 8-0 and 11-11 years of age. The hearing impaired/deaf children had no other significant handicaps, acquired their impairment before age 12 and had been educated using oral communication.

Reliability

Test-retest reliability for prompted responses based on retesting 26 normal-hearing children and 20 hearing-impaired children was .82 and .96, respectively; imitated response correlations were .81 and .95, respectively.

Validity

Concurrent validity was indicated by comparing results of the GAEL-C with the Carrow Elicited Language Inventory (CELI); correlation coefficients were .81 for prompted responses and .83 for imitated responses.

Special Alerts/Comments

The GAEL-C is designed to evaluate a student's acquisition of specific grammatical structures through prompted verbal responses. When appropriately completed, the test provides detailed analyses which may be useful for developing goals for IEPs and making instructional planning decisions. Limited standardization and technical development information preclude the use of results for making performance level decisions.

Due to lengthy test procedures, the student may tire easily and target sentences may not be obtained; therefore, results may not be indicative of a language problem. Also, scoring is somewhat difficult and time-consuming.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously. Scores should not be reported for bilingual and /or LEP students.

Reference

Manual

Publisher

Central Institute for the Deaf

Title	<u>Language Proficiency Tests (LPT)</u>	Grade	<u>Secondary-Postsecondary</u>
Author	<u>T. Geraro & G. Weinstock</u>	Time	<u>2 hours</u>
Copyright Date	<u>1981</u>	Type	<u>CRT</u>

Intended Purpose of Test

The LPT is designed to assess the level of language functioning of ESL and learning handicapped high school students and adults with low-level language skills.

Description

The LPT, consisting of individual test booklets with a multiple choice format and short answer section, contains the following subtests:

<u>Subtest</u>	<u>Contents</u>
<u>Aural/Oral</u>	
Commands	Oral commands which must be carried out by the student (optional for native English students)
Short Answer	Simple questions about home and school (optional for native English students)
Comprehension	Short passages read to student followed by oral questions
<u>Reading</u>	
Vocabulary	Choose correct words to complete given sentences
Comprehension	Reading passages ranging in difficulty from first through sixth grades, followed by multiple choice questions
<u>Writing</u>	
Grammar	Select the correct word or phrases to complete given sentences (including verbs, pronouns, plurals and comparisons)
Sentence Response	Brief questions to which the student must respond in writing
Paragraph Response	Write a paragraph about one of three given topics

Translation
(optional and advanced)

Translation by the ESL student of
his/her paragraph into his/her
native language

Test Administration

All subtests may be administered by teachers or other professionals on a group basis except for Commands and Short Answer which must be individually administered. No special training or materials are required and the test may be conducted in several brief sessions.

Subtests are scored separately; raw scores are multiplied to obtain adjusted scores which indicate the proficiency level the student has attained. An adjusted score of less than 50% indicates no proficiency, 50 to 69% - elementary placement, 70 to 84% - intermediate placement, and 85 to 100% indicates advanced placement.

Technical Information

The LPT is an outgrowth of the need to develop an instrument to appropriately assess adults whose language difficulties prevent success in academic or vocational areas.

The manual offers information regarding reliability and validity for only the Comprehension subtest. Test-retest reliability based on retesting 46 high school students at a two week interval was .87. Concurrent validity was demonstrated by comparing LPT scores with those obtained on the Barnet-Loft Multiple Skills Series, yielding a coefficient of .77.

Reading levels for the Reading Section were determined by use of the Fry Readability Formula.

Effective Use/Comments

The LPT is designed to assess the level of language functioning of ESL and learning handicapped high school students and adults with low-level skills. The information obtained is insufficient for determining performance levels in language proficiency. It is sufficient for making evaluation of student progress and instructional planning decisions and for developing long- and shortterm objectives on the IEP in the area of communication and language.

The test includes an optional translation for the ESL student which is "useful in assessing sophistication of vocabulary and syntactical competency"; it also "allows the language skills of students from many different backgrounds and cultures to be accurately evaluated" (Manual, p. 8).

Therefore, the LPT is recommended for use with linguistically and culturally diverse students and LEP students.

Reference

Manual

Publisher

Academic Therapy Publications

Title	<u>Lindamood Auditory Conceptualization Test (LAC)</u>	Age	<u>Pre- K to Adult</u>
Author	<u>C.H. Lindamood & P.C. Lindamood</u>	Time	<u>Untimed</u>
Copyright Date	<u>1971</u>	Type	<u>CRT</u>

Intended Purpose of Test

The LAC is designed to measure auditory perception of two related abilities: the ability to discriminate among speech sounds and the ability to perceive the number and order of sounds within a spoken pattern.

Description

The LAC consists of 18 colored wooden blocks (3 each of red, yellow, green, white, blue, black), Individual Record Sheets and a manual with instructions. Each colored block is used to represent an individual sound. The student places the blocks in a row from left to right to indicate or track the order of sounds in sequences. The student may choose any colors to represent any sounds as long as the block pattern is correct.

The test is divided into two categories. Category I includes isolated sounds in sequence, while Category II includes sounds within a syllable pattern. Two equivalent forms of the test are available.

Test Administration

The test is individually administered and the examiner requires no special training. A record is provided so that the examiner may be certain he/she is using correct pronunciation, but it is not intended to be used with the student. No additional materials are needed.

Technical Information

This test is an outgrowth of experience and observation of individuals with severe reading, spelling, or speech disorders. The underlying assumption is that if a person can perceive individual sounds, he/she may not necessarily be able to perceive these sounds in syllables.

Test-retest reliability based on retesting 52 students was .96. Predictive validity was determined by comparing LAC test results with scores on the WRAT in the areas of Reading and Spelling. The sample consisted of 660 students, grades K-12 in Monterey, California. A full range of socioeconomic levels and ethnic groups were represented. Correlations ranged from .66 to .81 with an average of .73. In a second sample of 52 students, correlations ranged from .72 to .78.

Effective Use/Comments

The LAC may be useful for obtaining descriptive information about the student's auditory perception. Lack of technical information precludes using the results for making screening, performance level, evaluation of student progress, and instructional planning decisions.

In order to administer the test, the student must be able to judge sameness and difference, number concepts to four, left-right progression, and first/last concepts. The LAC provides a precheck to test for these concepts.

The use of colored blocks was meant to eliminate complicating factors such as the need for knowledge of sound/symbol associations. However, difficulties may arise in the manipulation of the colored blocks. Inappropriate use of the blocks would result in invalid test scores.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

Reference

Manual

Publisher

DLM Teaching Resources

Title	<u>The Northwestern Syntax Screening Test (NSST)</u>	Age	<u>3 - 8 years</u>
		Time	<u>15 - 25 minutes</u>
Author	<u>L. Lee</u>	Type	<u>NRT</u>
Copyright Date	<u>1971 (Revised)</u>		

Intended Purpose of Test

The NSST is designed for screening the syntactic and morphological development of Standard English speakers. It is not intended as a general language development measure.

Description

The NSST is an individually administered device. It attempts to identify students who may have problems in receptive or expressive syntax or morphology. The NSST assesses: verb tenses, subject-verb agreement, question-statement difference, pronouns, plurals, possessives, negatives, prepositions, and Wh questions.

Norm Data

The NSST was standardized on 344 children, aged 3 years to 7 years, 11 months. These children were from middle and upper middle income families using standard American dialect. The sample was limited in size, economic class, and geographic region, and has been criticized as such. Applicability of these norms to other populations is questionable. The author has expressed the need for users to establish their own local norms.

Reliability

No reliability data are provided in the manual.

Validity

No data regarding the validity of the NSST are provided. The test relies on its content validity when used as instructed.

Special Alerts/Comments

The NSST provides information on syntactic and morphological development which may be useful in making screening decisions. Inadequate technical information precludes using test results for performance level instructional planning decisions.

It should not be used as a measure of general language skill.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

References

Buros, 1985
Salvia & Ysselayke, 1985

Publisher

Northwestern University Press

Title	<u>Peabody Picture Vocabulary</u> <u>Test - Revised (PPVT-R)</u>	Age	<u>2-1/2 to 40 years</u>
Author	<u>L.M. Dunn</u>	Time	<u>Untimed (10-20 minutes)</u>
Copyright Date	<u>1981</u>	Type	<u>NAT</u>

Intended Purpose of Test

The PPVT-R provides a measure of receptive vocabulary for standard American English Speakers.

Description

The PPVT-R is a revision of the 1959 edition of the PPVT and, while similarly available in two parallel forms, two-thirds of the items in the revised edition are new. Each item consists of four pictures presented in easel format. It is individually administered; the examiner reads a word for each plate and the student points to the picture which best represents the word.

Norm Data

The standardization sample consisted of 4,200 students between the ages of 2 years 6 months through 18 years 11 months, plus 828 adults from 19 to 40 years of age. Females and males were equally represented in both samples. Four geographic regions were used, excluding Alaska and Hawaii. The ethnic composition was similar to that of the total U.S. population. Subjects were selected from cities and suburbs; a cross-section of socioeconomic groups in the U.S. was included.

Reliability

Two types of reliability data were reported; measures of internal consistency ranged from .61 to .88; test-retest reliabilities ranged from .73 to .91. Alternate form reliability ranged from .65 to .89 (Bracken, McCallum, & Prasse, In Press).

Validity

The correlations of PPVT-R standard scores and the g score of several cognitive measures indicated that there is insufficient shared variance to justify using the PPVT as a measure of intelligence.

Special Alerts/Comments

The PPVT-R provides a measure of receptive vocabulary for standard American English speakers. This may be useful for screening decisions, providing descriptive information about a student's receptive language, and developing longterm objectives on the IEP in the area of communication and language.

Given the limited sample of behaviors and the reported validity information, a PPVT-R score may NOT be used as a measure of intellectual/cognitive functioning. The PPVT-R may be useful with a broad spectrum of the special education population since it requires no verbal response. Results of the PPVT-R are insufficient for making instructional planning and evaluation of student progress decisions.

Results with linguistically and culturally diverse students should be interpreted cautiously. It is not recommended for use with LEP students.

References

Publisher

Manual

American Guidance Service, Inc.

Salvia & Ysseldyke, 1985

Sattler, 1982

Bracken, McCallum, & Prasse, In Press

Title	<u>Receptive Language Inventory (RLI)</u>	Grade	<u>Preschool or Language Delayed</u>
Author	<u>M.L. Bowen, and J.C. Chalfant</u>	Time	<u>30 minutes</u>
Copyright Date	<u>1981</u>	Type	<u>CRT</u>

Intended Purpose of Test

The RLI is designed to assess levels of receptive language usage as well as overall language comprehension. It is intended for use with individuals who do not talk or who exhibit discrepancies in language development and/or disabilities in language acquisition and use. It may be used to assess language abilities of preschool, preschool handicapped, moderately and severely mentally retarded, language deficient, learning disabled, hearing impaired, and bilingual students as well as older students with delayed expressive language or language comprehension difficulty.

Description

The RLI consists of eight subtests in the following areas:

1. Attending establishing the relationship between stimulus and purposeful reaction (e.g., looking at a ball and showing what to do with it)
2. Chaining/Motor Imitation attending to and imitating a motor activity or sequence (e.g., watching the examiner jump and then jumping)
3. Chaining/Vocal Imitation attending to and imitating a vocal chain of two or more speech sounds (e.g., hearing examiner say "dip" and then repeating it)
4. Chaining/Objects attending to a request to perform a particular motor chain and performing the chain with an appropriate stimulus object (e.g., selecting a cup from a set of three objects and producing appropriate actions for drinking)
5. Chaining/Actions attending to a request to perform a particular motor chain and performing it (e.g., clapping hands upon request)
6. Functional discriminations identifying a particular concept by selecting one picture out of four (e.g., pointing to the picture of mother)
7. Labeling naming a picture object

8. Functional Concepts

identifying a particular concept by classification by selecting one picture out of four (e.g., pointing to the steak when asked to point to meat)

The complete RLI Kit contains an examiner's manual, individual response record booklets, spiral bound picture stimuli books, and small common objects. Rules for scoring differ among subtests. However, a score of 30% or above is required for mastery in each subtest area.

Test Administration

The RLI is individually administered by a teacher or trained paraprofessional. No special materials are required.

Technical Information

The RLI was piloted on 400 mentally/physically handicapped, learning/language disabled, and nonhandicapped students, ages 2.6 to 9.6.

Test retest correlations for 160 students were moderate to strong for all subtests (r ranged between .77 and .94) except for the Attending ($r=.34$). Content validity was developed from a review of item pools from participating classroom teachers.

Effective Use/Comments

The RLI provides information on receptive ability which is sufficient for developing short- and longterm goals for the IEP in the area of communication arts. Given the lack of standardization, the results are not sufficient for making performance level decisions. The language areas covered are prerequisite skills for those identified in the MTEs. While the instructional objectives thus developed are specific, the manual does not provide strategies for classroom instruction.

The authors recommended use of the RLI with bilingual students. It may be appropriate for students from culturally and linguistically diverse backgrounds provided that the language provisions are made by the interviewer.

Reference

Manual

Publisher

C.C. Publications Inc.

Title	<u>Test for Auditory Comprehension of Language-Revised Edition (TACL-R)</u>	Age	<u>3 to 10 years</u>
Author	<u>E. Carrow-Woolfolk</u>	Time	<u>Untimed (10 - 20 minutes)</u>
Copyright	<u>1985</u>	Type	<u>NRT</u>

Intended Purpose of Test

The TACL-R is designed to measure the auditory comprehension of language and determines the content areas of language difficulty.

Description

The TACL-R consists of 120 plates of line drawings illustrating receptive, syntactic, semantic, and morphological content. It is individually administered in three sections of 40 items. The section on "Word Classes and Relations" tests nouns, verbs, adjectives and adverbs. The section on "Grammatical Morphemes" tests prepositions, noun-verb agreement, noun number and case, verb number and tense (all in sentence context). The "Elaborated Sentence Constructions" includes embedded sentences, and partially and completely connected sentences. No verbalization is required as the student responds by pointing to one of three pictures per plate. Unlike the 1973 TACL edition, no Spanish version is available.

Norm Data

The standardization sample was based on 1,003 individuals between the ages of three and ten. The sample was stratified by family occupation, ethnic origin, age, sex, community size and geographical distribution and was consistent with the 1981 U.S. Census.

Results reported as percentile ranks, age equivalent scores, and standard scores by age level and grade level (K - 4).

Reliability

The three measures of reliability examined (standard error of measurement, split-half and test-retest) all show that the TACL-R produces highly reliable scores for all three sections as well as the total score; test-retest reliability ranged from .89 to .95.

Validity

Construct validity was developed through an indication of the relationship of scores to language development. Content validity was established through comparison studies between "normal" and communicatively impaired individuals; the performance differences were larger for the group with language comprehension as a primary disorder than for the group with difficulties in oral language expression. Overall, the performance of individuals classified as having articulation as their primary disorder did

not differ from the norming sample. The results of the articulation group differed from results used in the validity of the 1973 TACL. The author suggested that perhaps additional language problems existed with the 1973 group and that this bears further investigation.

Special Alerts/Comments

The TACL-K is designed to measure the auditory comprehension of language and determines the content areas of language difficulty which may be useful for making screening, performance level, evaluation of students progress and instructional planning decisions and for developing short- and longterm objectives on the IEP in the area of communication and language.

It is not recommended for use with linguistically and culturally diverse students and LEP students.

References

Manual

Publisher

DLM Teaching Resources

Title	<u>Test of Adolescent Language (TOAL)</u>	Age	<u>11 to 18 years</u>
Author	<u>D.O. Hammill, V.L. Brown, S.C. Larsen, & J.L. Wiedernolt</u>	Time	<u>Untimed (5-15 minutes per subtest)</u>
Copyright Date	<u>1982</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The TOAL is designed to assess the language proficiency of adolescents with respect to listening, speaking, reading, and writing.

Description

The TOAL consists of eight subtests which estimate a student's vocabulary and grammar proficiency in listening, speaking, reading, and writing. Only the speaking subtests must be administered individually.

Norm Data

The TOAL was normed on a sample of 2,723 students, ages 11 to 18, from 17 states and three Canadian provinces. The sex and residence characteristics were found to be nearly identical to those of the national population reported in the 1977 census. No handicapped students were included.

Results are reported as raw scores, scaled scores, composite scores, and language quotients.

Reliability

Test-retest reliability based on retesting 52 students from a parochial school in Kansas City, Missouri ranged from .82 to .98. Interrater reliability ranged from .87 to .98.

Validity

Criterion-related validity, was based upon comparisons of the TOAL with the PPVT, the DTLA, and the TWL, and subtests from CTBS; TWL; correlations ranged from .59 to .83.

Content validity was indicated through test construction and item selection. Construct validity was demonstrated by age and group differentiation, the relationship to intelligence tests, and subtest interrelationships. Subtest means increased with student's age. The correlation of the TOAL with the California Short-Form Test of Academic Aptitude was .79. Correlation coefficients with respect to subtest interrelationships ranged from .32 to .75.

Effective Use/Comments

The TOAL is designed to assess the language proficiency of adolescents with respect to listening, speaking, reading, and writing. The findings may be useful for screening, evaluation of progress, and instructional planning decisions and for developing short- and longterm objectives on the IEP in the area of communication and language. Given its limited standardization sample it may not be used to determine a level of performance.

While objectives are specific, intervention strategies were not provided. The authors cautioned the value of attempting remediation in specific subtest areas but listed resources for further assessment and programing. Composite scores must not be used as scores of mental ability.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously. Scores may not be reported for bilingual and/or LEP students.

References

Manual
Salvia & Ysseldyke, 1985

Publisher

PRO-ED

Title	<u>Test of Early Language Development (TELD)</u>	Age	<u>3 to 8 years</u>
Author	<u>W.P. Hresko, D.K. Reid, & D.D. Hammell</u>	Time	<u>Untimed</u>
Copyright Date	<u>1981</u>	Type	<u>NAT</u>

Intended Purpose of Test

The TELD is designed to identify students who are significantly behind their peers in the development of language, to document progress in language, and to plan language instruction.

Description

The features of language measured by the TELD are phonology, morphology, syntax, and semantics in both the expressive and receptive modes. The test is individually administered, and consists of 38 items. Each item assesses one mode and one feature. Students begin the test at an age appropriate starting point. The pass/fail criterion for each item is clearly delineated in the test manual.

Norm Data

The TELD was standardized on the test performance of 1,184 students in eleven states and one Canadian province. The sample closely reflected U.S. population as reported in the 1979 Statistical Abstract of the United States.

Reliability

Both internal consistency and test-retest reliability coefficients were reported as .90 for all age levels.

Validity

Content validity was indicated through test development procedures. TELD scores were correlated with a variety of achievement, readiness, intelligence, and language tests which resulted in satisfactory criterion-related validity and construct validity.

Special Alerts/Comments

The TELD provides a gross measure of a student's language development which may be useful for making screening, evaluation of student progress, and instructional planning decisions. Given its limited standardization sample, it may not be used to determine a level of performance.

Reasonable cautions should be used when interpreting TELD scores. It should be recalled that the TELD measures many but by no means all of the abilities that may be involved in early language.

The TELD is not appropriate for use with bilingual and/or LEP students.

Reference

Manual

Publisher

PRO-ED

Title	<u>Test of Language Development</u> <u>Intermediate (TOLD-I)</u>	Age	<u>8 1/2 to 13 years</u>
Author	<u>P. L. Newcomer & D. D. Hammill</u>	Time	<u>Untimed (40-50 minutes)</u>
Copyright Date	<u>1982</u>	Type	<u>NKT</u>

Intended Purpose of Test

The TOLD-I is designed to identify broad areas of a student's language difficulties and serves as the basis for planning an extensive program of criterion and diagnostic teaching.

Description

Individually administered, the TOLD-I has five principal subtests and two supplemental subtests: Semantics (picture vocabulary and oral vocabulary); Syntax (grammatical understanding, sentence imitation, and grammatical completion). The resulting performance provides a Language Age (LA) for each subtest and a Language Quotient (LQ) for the total test. Specific cautions and limitations of test results are explicitly detailed in the manual.

Norm Data

The TOLD-I was standardized on an unselected sample of 1,014 student's residing in 15 states. The standardization sample compared very closely to the 1970 census with respect to sex, area of residence (urban or rural), race, geographic area, and occupation of parents.

Reliability

Types of reliability reported include internal consistency and test-retest. In general, the coefficients were found to be greater than .90 at most age levels. The standard errors of measurement varied considerably, reflecting the reliability and standard deviations of subtest and total scores at each level.

Validity

The authors reported five types of validity: content, item, concurrent, construct, and diagnostic. Based on correlations with other specific measures of articulation, syntax, and vocabulary (e.g., PPVT, Northwestern Syntax Test), concurrent validity coefficients were in the .70's and .80's. With respect to diagnostic validity, the test seems to differentiate between students of normal and impaired language.

Special Alerts/Comments

The TOLD-I provides a gross measure of a student's language development which may be useful for making screening decisions. It is insufficient for determining a student's performance level in language. As stated by the authors, "it does not constitute a suitable base for planning instructional programs."

While the authors provide specific cautions regarding the test, the use of the Linguistic Quotient (LQ) is not adequately explained in terms of educational usefulness.

The word Discrimination and Word Articulation subtests should not be given to students older than six. It is most appropriate for younger children, and is easy to administer.

It is not appropriate for use with LEP students.

References

Manual:
McLoughlin & Lewis, 1981

Publisher

PRO-ED

Title	<u>Test of Language Development - Primary (TOLD-P)</u>	Age	<u>4 to 9 years</u>
Author	<u>P. L. Newcomer & D. D. Hammill</u>	Time	<u>Untimed (40-50 minutes)</u>
Copyright Date	<u>1982</u>	Type	<u>NRT</u>

Intended Purpose of Test

The TOLD-P is designed to identify broad areas of a child's language difficulties and serves as the basis for planning an extensive program of criterion and diagnostic teaching.

Description

Individually administered, the TOLD-P has five principal subtests which provide information on a student's syntax and semantics. Phonology is supplemental. The resulting performance provides a Language Age (LA) for each subtest and a Language Quotient (LQ) for the total test.

Norm Data

The TOLD-P was standardized on 1,836 children residing in 19 states and one Canadian province. The sample was stratified by sex, area of residence (urban or rural), race, geographic area, and occupation of parents and was consistent with the 1970 U.S. Census.

Reliability

Internal consistency, reported in five independent studies, was not always adequate. Test-retest reliability, based on retesting 21 students over a 5-day interval was high (above .86). The standard error of measurement for subtests ranged from 0.7 to 3.0; composites ranged from 2.1 to 8.6.

Validity

The authors reported three types of validity: content, criterion, and construct. Content validity was indicated through consistency with popular taxonomies of language. Construct validity was demonstrated by the TOLD-P's relationship to similar scales. Based on correlations with other specific measures of articulation, syntax, and vocabulary (e.g., PPVT, WISC Vocabulary, DTLA-Related syllables), concurrent validity coefficients ranged from .49 to .86.

Special Alerts/Comments

The TOLD-P provides a gross measure of a student's language development which may be useful for making screening decisions. It is insufficient for determining a student's performance level in language. As stated by the authors, "results constitute neither a suitable nor sufficient basis for planning instructional programs" and "should not serve as the basis for an Individualized Education plan" (p.23).

The word Discrimination and word Articulation subtests should not be given to students older than six years.

The Spoken Language Quotient (SLQ), derived by pooling the standard scores of the five principal subtests, must not be used as a measure of student IQ.

Results with linguistically and culturally diverse students should be interpreted cautiously. It is not recommended for use with LEP students.

References

Manual

Publisher

PRO-ED

Title	<u>Test of Syntactic Abilities (TSA)</u>	Age	<u>10 to 19 years</u>
Author	<u>S.P. Quigley, M.W. Steinkamp, U.J. Power, & B.W. Jones</u>	Time	<u>30 min. (each diagram test 1 hour Screening Test)</u>
Copyright Date	<u>1978</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The TSA is designed to identify language difficulties with syntactic structures for deaf students.

Description

The TSA consists of: 1) the Screening Test containing 120 multiple-choice items and 2) the Diagnostic Battery of 20 individual diagnostic tests, each containing 70 multiple-choice items; each test may be used as an independent assessment instrument. The Screening Test and Diagnostic Battery each include nine major syntactic structures of standard English.

Test administration and scoring are simple and require no special training. Parallel forms are available for both the Screening Test and Diagnostic Battery.

Norm Data

The standardization sample consisted of approximately 450 deaf students, ages 10 to 19 years, from 18 residential and day programs in nine regions of the United States. Males and females were equally represented. Ethnic composition was not addressed. The sample was proportional to the number and distribution of deaf students established by the U.S. Census. The students had no apparent disability other than hearing impairment.

Results are reported as raw scores, percentile ranks, and age equivalents.

Reliability

Internal consistency for each of the 20 diagnostic tests was above .93. Test-retest reliability based on retesting 54 students over a 6 month interval ranged from .62 to .87. Intercorrelations among the 20 tests ranged from .56 to .88.

Validity

Content validity was developed through ten years of research on linguistic structures considered most important for deaf children to learn. Efforts were made to maintain reading vocabulary within the comprehension level of deaf children and to minimize sexual and cultural bias. Concurrent validity was demonstrated through correlations between TSA structure scores and scores on language-related subtests of the SAT. Point biserial correlation was greater than .40 for 88% of the items in the battery.

Special Alerts/Comments

The TSA is designed to identify language difficulties with syntactic structures for deaf students. The information obtained is useful for screening purposes, evaluation of student progress, making instructional planning decisions, and developing short- and longterm objectives on the IEP in the area of Communication and Language.

Results with linguistically and culturally diverse students should be interpreted cautiously. It is not recommended for use with LEP students.

Reference

Manual

Publisher

Dormac, Inc.

Title	<u>Woodcock Language Proficiency Battery (English) (WLPB)</u>	Age	<u>3 - 80 years</u>
Author	<u>R. Woodcock</u>	Time	<u>Untimed (15-45 mins.)</u>
Copyright	<u>1980</u>	Type	<u>NRT</u>

Intended Purpose of Test

The purpose of the WLPB is to measure proficiency in English of oral language ability, aspects of reading ability, and aspects of written language ability. When the English form is used in conjunction with another language form, measures of relative language proficiency may be obtained.

Description

There are three subtests of oral language ability including picture vocabulary, antonyms, synonyms, and analogies; three subtests of reading ability including letter-word identification, word attack, passage comprehension; and two subtests of written English which assess skills in punctuation, capitalization, spelling, and usage. The test is contained in a flip-page easel. A response booklet is used for recording purposes, summarizing results, and interpreting test performance. Four profiles are included on the Response Booklet: Percentile Rank Profile, Subtest Profile, Instructional Implications Profile, Relative Performance Index. This test has some additional clusters of language subtests which make it different from the W-J Battery.

Norm Data

The test was normed on a stratified random sample of 4,732 subjects in 40 communities in the U.S. Subjects ranged in age from 3 to 80 years. Data are identical to those reported for the W-J Battery.

Reliability

Data are identical to those reported for the W-J Battery.

Validity

Validity is identical to the W-J Battery.

Special Alerts/Comments

The WLPB provides a measure of language proficiency in English with respect to oral language ability and aspects of reading and written language ability. Results may be useful for obtaining supplementary, descriptive information about a student's language proficiency level in English. Given the validity data and the relative recency of publication, the results of the WLPB are insufficient for determining the level of language proficiency in English and for making instructional planning decisions.

Scoring procedures are complex and time consuming and require practice. It is recommended that a small calculator be used for accuracy in test scoring.

A Spanish version of the WLPB is available and reviewed in the LEP section of this Guide.

Reference

Publisher

Manual

DLM Teaching Resources

PERCEPTUAL AND MOTOR FUNCTIONING

These tests measure skills in perception (visual, auditory), integration of perceptual abilities, and fine and gross motor functions. Some tests provide an indication of possible neurological impairment.

Caution should be taken when using these tests with physically handicapped and/or linguistically diverse students.

From a developmental perspective, they are primarily intended to provide information about the early stages of development.

Title	<u>Auditory Discrimination Test (ADT)</u>	Age	<u>5 - 8 years</u>
Author	<u>J.M. Wepman</u>	Time	<u>Untimed (Approx. 10 mins.)</u>
Copyright Date	<u>1973</u>	Type	<u>NRT</u>

Intended Purpose of Test

The ADT is designed to measure auditory discrimination abilities of young children.

Description

Individually administered, the ADT consists of a list of forty pairs of words, ten identical and thirty which differ from each other by one phoneme. The examiner reads each word pair to the student who must indicate whether the words are the same or different. Two forms of the test are available.

Norm Data

Data concerning the normative sample were not provided.

Reliability

Test-retest reliability coefficients were reported as .91 and .95. Alternate form reliability was reported as .92. Time intervals and sample identification were omitted.

Validity

No evidence of content or criterion validity was presented in the manual. A number of studies were conducted on the construct validity of the ADT (see Salvia and Ysseldyke for a complete review). Correlations with the MAT yielded coefficients of .235 and .348.

Special Alerts/Comments

The ADT provides an indication of auditory discrimination abilities. Given its technical limitations, the results of the ADT may be useful for obtaining supplementary, descriptive information about a student's auditory discrimination. However, it is insufficient for determining a performance level in auditory discrimination ability. As an informal tool, it is limited to providing gross screening information. Since it measures only a small aspect of auditory discrimination ability, inferences regarding a student's total auditory discrimination ability cannot be made. The relationship of the performance on the ADT and academic learning and speech ability is not established. Use with LEP students and linguistically and culturally diverse students may not be appropriate.

Reference

Manual
Salvia & Ysseldyke, 1985

Publisher

Language Research Associates

Title The Bender Visual Motor Gestalt
Test (Bender Gestalt)

Age 5 - 11 years

Author L. Bender & E. Koppitz (scoring for
young children)

Time Untimed

Copyright Date 1963, 1975 (Koppitz)

Type NRT

Intended Purpose of Test

The Bender Gestalt was originally developed by Bender in 1938. Bender used the test to differentiate brain injured adults and to detect signs of emotional disturbances. Over the years a number of scoring systems have been developed, the most common of which is the system developed by Koppitz. The present discussion of the Bender Gestalt is based solely on the Koppitz scoring system.

The purpose of the Koppitz scoring system is to assess perceptual maturity, possible neurological impairment, and emotional adjustment of young children.

Description

The Koppitz system scores the student copy of each design in terms of the absence or presence of specific errors: distortion of shape, perseveration, integration, and rotation. The student's score is the sum total of errors. The Koppitz system is only available in book form (Koppitz, 1963, 1975). The Bender Gestalt is untimed and is generally administered individually.

Norm Data

The test was normed in 1963 on 1,104 students from 46 classes in 12 public schools with 98% of the sample being white. It was renormed in 1974 on 975 students between 5 and 11 years of age. Racial balance was obtained and community size was varied. Geographical balance was not obtained and social/emotional status was not indicated.

Reliability

According to 23 studies (Koppitz, 1975), interscorer reliability ranged from .79 to .99 with 81% above .89. Test-retest reliability with normal elementary students ranged from .50 to .90 with a mean of .71. However, Salvia and Ysseldyke (1985) indicated that five of the nine reliability studies reported by Koppitz were conducted with kindergarten pupils only. Out of the 25 reported reliability coefficients, only one exceeded .90.

Validity

Koppitz has reported studies validating the test as an aid, when used with other measures, in the diagnosis of minimal brain-dysfunction and emotional disturbance. According to Salvia and Ysseldyke (1985) "the construct of visual-motor perception is never adequately defined in either Koppitz manual" (p. 270).

Special Alerts/Comments

The Bender Gestalt provides an indication of a young student's perceptual maturity, possible neurological impairment, and emotional adjustment. Because of unestablished reliability and validity, it is insufficient for making performance level, instructional planning, and/or evaluation of student progress decisions. The test results are best used in combination with other medical and behavioral data as supplementary, descriptive information.

Koppitz (1975, p. 70) indicated the test cannot be used to predict academic performance of individual students nor is there evidence that it aids in instructional planning. Inferences regarding the total domain of visual perception cannot be made by the use of the test since it is solely a measure of the student's ability to copy geometric designs. In order to score and interpret the results of the Bender Gestalt, training is required.

Although there is no evidence limiting its use with the LEP students, recent research (Sattler & Gwynne, 1982) suggested that performance on the Bender Gestalt may be affected by experience with other visual-motor tasks (e.g., coloring, tracing) and use of verbal mediation strategies.

References

Publisher

Manual
Anastasi, 1976
Mutt & Briskin, 1960
Koppitz 1963, 1975
Salvia & Ysseldyke, 1985
Sattler & Gwynne, 1982

Grune & Stratton

Title Bruininks-Oseretsky Test of Age 4.6-14.5 years
Motor Proficiency (Bruininks-Oseretsky)
Author R. Bruininks Time 45-60 minutes
(short form 15 minutes)
Copyright Date 1978 Type NRT

Intended Purpose of Test

The Bruininks-Oseretsky test is designed to assess motor skill ability and development.

Description

This individually administered test is made up of 46 items grouped into eight subtests: running, speed and agility, balance, bilateral coordination, strength, upper limb speed, dexterity. Separate scores of gross and fine motor skills plus an overall index of motor proficiency are obtained. The test provides subtest scores, a Gross Motor Composite score, a Fine Motor Composite score, and a Battery Composite score.

Norm Data

The Bruininks-Oseretsky was standardized on a population of 765 boys and girls selected from various schools, day-care centers, nursery schools, and kindergartens in the United States and Canada. A stratified sampling procedure, based on the 1970 U.S. Census, was used to select the students. Stratification variables included age, sex, race, community size, and geographical region.

Results are reported as standard scores, percentile numbers and stanine are provided.

Reliability

Internal consistency was demonstrated by correlating item scores with subtest and total test scores. These correlations were generally between .70 and .80. Test-retest reliabilities ranged from .86 to .89 for the Battery Composite, and from .68 to .88 for the Fine and Gross Motor Composites. Average standard errors of measurement were 4.0 for the Battery Composite, 4.6 for the Gross Motor Composite, and 4.7 for the Fine Motor Composite. Low reliability coefficients were reported for some subtests.

Validity

Validity was determined by correlating test score and age; coefficients ranged from .57 to .86, with a median of .78. A factor analysis, performed on the standardized sample, provided limited support for grouping the items into various subtests. Five factors were found, with one factor (general motor ability) accounting for approximately 70% of the total common factor item variance. The remaining factors represented specific factors associated with the various subtests. Most of the items measuring fine motor skills did not cluster together on clearly identifiable factors as did the gross motor items.

Special Alerts/Comments

The Bruininks-Oseretsky provides information about a student's performance level in motor skill abilities and development. The information from the Bruininks-Oseretsky may also be useful for making instructional planning and evaluation of student progress decisions with respect to motor ability.

Caution should be used in interpreting findings based on individual subtests. It is not appropriate for severely physically impaired students. The English test instructions are inappropriate for the LEP students.

References

Manual
McLoughlin & Lewis, 1981

Publisher

American Guidance Service,
Inc.

Title	<u>Developmental Test of Visual-Motor Integration (VMI)</u>	Age	<u>2 - 19 years</u>
Author	<u>K. Beery & N. Buktenica</u>	Time	<u>Untimed</u>
Copyright Date	<u>1967, 1982</u>	Type	<u>NRT</u>

Intended Purpose of Test

VMI is designed to measure the integration of visual perception and motor coordination.

Description

VMI consists of 24 geometric designs that are copied by a student in a test booklet. The geometric forms are arranged in order of difficulty. It may be administered to individuals or to a group by a classroom teacher.

Norm Data

A description of the norm sample is not provided in the manual, but is contained in a separate monograph. A total of 1,039 middle-class students from Illinois were included in the standardization. While six demographic characteristics were reported, they were not included in the manual.

The test was renormed in 1981 based on the testing of 3,090 students. Information on the sample's ethnic composition, income level, residential area, and sex were reported; geographic area data were omitted.

Percentiles and standard scores were reported by age.

Reliability

Three kinds of reliability were reported: interscorer reliability, test-retest reliability, and internal consistency. Internal consistency obtained from four studies ranged from .66 to .93. Test-retest reliability for 171 students over a 2-week interval was .83 for the boys and .97 for the girls; in five other studies, reliability coefficients ranged from .63 to .92. Interscorer reliability obtained from these studies ranged between .58 and .99.

Validity

The content validity for the VMI was based on expert opinion on selection of items. According to Buros (1978), construct validity was questionable because visual motor integration is not adequately defined. Predictive validity has not been established; studies relating VMI and achievement showed mixed results.

Special Alerts/Comments

VMI provides an indication of the integration of visual perception and motor coordination which may be used to obtain supplementary, descriptive information about a young student's visual motor integration. However, as a measure of visual-motor integration, the behavior assessed is limited to reproduction of geometric designs. Inferences to the total visual motor perception domain cannot be made. Since the relationship between visual motor integration as measured by the VMI and academic achievement has not been established, inferences regarding academic performance are not warranted. At the upper end of the age range (10+), small gains in raw scores produce disproportionate gains in the age equivalent scores, leading to a high potential for diagnostic errors for students over the age of ten (Synder, Synder, & Massong, 1981).

There is no evidence regarding its use with linguistically and culturally diverse students and LEP students.

References

Publisher

Manual
Burros, 1978
Salvia & Ysseldyke, 1985
Synder, Synder, Massong, 1981

Follett Publishing Co.

Title	<u>Project Unique Physical Fitness Test</u> <u>(Project Unique)</u>	Age	<u>10 - 17 years</u>
Author	<u>J.P. Winnick & F.X. Short</u>	Time	<u>Two sessions</u> <u>(40 minutes each)</u>
Copyright Date	<u>1985</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

Project Unique is designed to measure the physical fitness of students with orthopedic disabilities (amputations, congenital anomalies, cerebral palsy, spinal neuromuscular conditions) and/or sensory impairments (auditory, visual). It may also be used with nonhandicapped children.

Description

The test may be administered in the gymnasium or outdoors. It measures fat % (triceps, subscapular), arm and leg strength (grip strength, broad jump), power and speed (50 yd. dash), power and abdominal strength (sit-ups), power and upperbody strength (arm hang, softball throw), and abdominal flexibility (sit and reach). Test items vary based upon the student's disability.

Norm Data

Standardization was based on testing 3,914 handicapped and nonhandicapped students, ages 10 through 17, from 23 states throughout the country. The sample was stratified according to age, gender, disability, school setting (institutional, noninstitutional), community size, and geographic region.

Reliability

Test reliability, as documented by Daquila (1982), was adequate.

Validity

Content validity was developed as a result of item selection and test construction. Construct validity, indicated through factor analysis, resulted in factor loadings of .40 and above on at least four out of six different factor solutions.

Special Alerts/Comments

Project Unique is designed to measure the physical fitness of students with orthopedic disabilities and/or sensory impairments, as well as nonhandicapped students which may be useful for planning instruction, evaluating on adaptive physical education program, and developing short- and longterm objectives on the IEP in the area of physical development.

Inadequate technical information precludes using results for determining performance level.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

References

Manual
Daquila, 1982

Publisher

Human Kinetics Pub.

Title	<u>Revised Visual Retention Test (Benton)</u>	Age	<u>8 - 64 years</u>
Author	<u>A. Benton</u>	Time	<u>5 - 10 minutes</u>
Copyright Date	<u>1974</u>	Type	<u>NRT</u>

Intended Purpose of Test

The purpose of the Benton is to assess clinically visual perception, visual memory, and visuoconstructive abilities.

Description

There are three alternate forms of the test. Each form consists of ten designs in a spiral notebook, each design containing one or more figures. It may be administered in four different ways: exposure of the design for either five or ten seconds followed by immediate production from memory, design copied by subject while in view, design exposed for ten seconds and copied from memory after a delay of 15 seconds. Scoring is based on the number of correct designs and total number of errors.

Norm Data

Norms were presented for only three modes of administration; the fourth is not yet complete. Norms were given for children of ages eight through fourteen and for adults fifteen to sixty-four. Different norms were presented for different IQ levels. Description and analysis of the standardization sample in the manual were sparse; the user was referred to the American Documentation Institute for additional information.

Reliability

Interscorer reliability was reported as .95 for the total score and ranged from .75 to .98 for error categories. Coefficients of equivalence between the three forms ranged from .79 to .84.

Validity

Numerous studies were cited by Benton (1974) demonstrating that various types of brain injured subjects perform poorly as compared with nonbrain-damaged subjects on this test. However, no evidence was given that the test distinguishes between brain damage and perceptual disorders.

Special Alerts/Comments

The Revised Benton provides an indication of visual perception, visual memory, and visuoconstructive abilities. Because of the technical inadequacy of the Visual Retention Test, it is insufficient for making performance level, instructional planning, or evaluation of student progress decisions. It may be useful in providing supplementary, descriptive information with respect to visual perceptual memory and visual motor integration.

There is no evidence limiting its use with linguistically and culturally diverse students and LEP students.

Scores should not be reported for bilingual and/or LEP students.

References

Manual
Anastasi, 1976

Publisher

The Psychological Corp.

Title	<u>Test of Auditory Discrimination (TAD)</u>	Age	<u>3 - Adult</u>
Author	<u>R. W. Woodcock</u>	Time	<u>Untimed (10-15 minutes)</u>
Copyright Date	<u>1970</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

TAD is designed to identify and assess the listener's ability to distinguish among speech sounds in a quiet and a noisy atmosphere. Measures of auditory discrimination are obtained under ideal conditions and in the presence of background cafeteria noise.

Description

The test consists of the presentation of individual words on a cassette tape. The subject is asked to point to one of four pictures on a plate that best corresponds to the word. The test has three parts: (1) Training procedure where the subject is familiarized with the pictures and names given on the two subtests (2) Quiet subtest in which individual words are presented in the absence of noise (ideal condition), and (3) Noise subtest in which words are presented with background noise. This last subtest provides a measure of auditory discrimination under conditions similar to everyday life. TAD may be group or individually administered.

Norm Data

TAD was standardized on 745 individuals, 3 to 84 years old, living in New Jersey, Minnesota, and Tennessee. People with hearing loss were excluded from the sample. No information was given concerning the socioeconomic status or ethnicity of the sample.

Reliability

Split-half reliability was reported as .79 for the Quiet Subtest and .68 for the Noise Subtest. In a clinical sample of 242 children, ages 4 to 12, split-half reliability was .87 for the Quiet Subtest and .68 for the Noise Subtest. For a sample of 17 preschool speech-handicapped children, test-retest reliability coefficients of .87 for the Quiet Subtest and .81 for the Noise Subtests were obtained.

Validity

The author reported that the test has content, concurrent, and construct validity. This claim was based on the nature of the test and correlation with expert judgments. Low correlations were reported with other unspecified tests. Comparisons were also made between normal and clinical groups. Correlations with the S-B among 40 preschool disadvantaged children yielded coefficients of .60 for the Quiet Subtest and .52 for the Noise Subtest.

Special Alerts/Comments

TAD provides information about the student's ability to discriminate among speech sounds. Although technically it meets criteria as a norm referenced test, TAD may be useful as a criterion referenced measure. It may be useful in making screening decisions when used as part of a complete language assessment battery. As a criterion referenced tool, it may provide some information for instructional planning decisions. However, according to Banatyne (cited in Sattler, 1982), validity is limited as it fails to differentiate adequately among age groups.

It should be noted that the test has been revised and extended as part of the Goldman-Fristoe-Woodcock Auditory Skill Test.

Given the restricted sample and the demands of the test, it is not appropriate for use with linguistically diverse and LEP students.

References

Manual
Sattler, 1982

Publisher

American Guidance Service, Inc.

Title	<u>Test of Gross Motor Development</u> <u>(TGMD)</u>	Age	<u>3 - 10 years</u>
Author	<u>D.A. Ulrich</u>	Time	<u>Untimed (approximately</u> <u>20 - 30 minutes)</u>
Copyright Date	<u>1985</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The TGMD is designed to measure the level of gross motor development of children in preschool and early elementary grades.

Description

The TGMD consists of 12 gross motor skills which are grouped into two subtests: locomotion and object control. Skill sequence is stressed rather than the product of performance. The test is individually administered by a professional who requires a minimum amount of specific training.

Norm Data

The standardization sample consisted of 909 children residing in eight states - California, Georgia, Illinois, Kentucky, Maryland, Missouri, Pennsylvania and Wyoming. The sample was controlled for age, gender, race, community size and geographic region.

Reliability

Interscorer reliability for the locomotion and object control subtests was .53 and .81, respectively. Test-retest reliability was .53 for the locomotion subtest and .49 for object control.

Validity

Construct validity was indicated through factor analysis of the 12 gross motor skills. The principal components method of factor analysis with varimax rotation was used. Nine of the twelve skills had loadings on a single gross motor factor at or above .50. By incorporating the percent of variance criterion of 70% and the score test, it appeared that three factors could be extracted. The skills were highly related in that 75% of the common variance shared by all 12 skills was explained by these factors.

Special Alerts/Comments

The TGMD is designed to measure the level of gross motor development of children in preschool and early elementary grades which may be useful for making instructional planning decisions, evaluating student progress, and developing long- and shortterm objectives on the IEP in the area of physical development. It may also be useful for screening purposes. Because of limited standardization, it may not be useful for determining performance level.

The test is sensitive to formal instruction in fundamental gross motor development. Preschool teachers should feel confident that when reasonably effective instruction is implemented for a minimum of 20 weeks, student will show performer improvement reflected in increased test scores.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

References

Manual
Ulrich, 1985

Publisher

PRO-ED

Title	<u>Visual Aural Digit Span Test (VADS)</u>	Age	<u>6- 1/2 to 13</u>
Author	<u>E. Koppitz</u>	Time	<u>10 minutes</u>
Copyright Date	<u>1977</u>	Type	<u>NRT</u>

Intended Purpose of Test

VADS is designed to measure aural and visual short term memory and provide information about intersensory and intrasensory integration.

Description

VADS consists of three subtests: (1) the Aural-Oral subtest requires the oral repetition of orally presented digits, similar to the Digit Span subtest of the WISC and S-B, (2) the Visual-Oral subtest requires the student to repeat from memory visually presented digits; (3) the Aural-Written subtest requires the writing of digits that are visually presented. The score for each subtest is the longest digit sequence successfully completed.

VADS yields 11 different scores: three from the subtests, seven from the summation of subtest scores which provide intra- and intersensory integration scores, and a total score.

Norm Data

The standardization sample consisted of 810 students between 5 and 13 years of age who were living in New York, Ohio, California, Virginia, and Texas. Equal numbers of boys and girls were included at each age level. The sample was representative of socioeconomic and ethnic groups in the U.S. Students who were seriously handicapped were not included.

Raw scores were converted to grade and age percentile norms.

Reliability

Test-retest reliability was reported for two groups of students, 6 to 10 years and 11 and 12 years; correlations for both groups ranged between .70 and .95 for the 11 separate scores.

Validity

Correlations between VADS and WRAT scores among learning disabled 8- and 10-year-old students ranged from .29 to .65. Among other LD samples, correlations between the VADS total score and WISC IQ's were .45 (verbal), .39 (full scale), and .26 (performance); the latter coefficient was not significant. Koppitz (1977) reported that the VADS discriminated LD students from the general population in second, third, and fourth grades; however, there were no differences between average and above average students in the first through fifth grades.

Special Alerts/Comments

VADS provides an indication of aural and visual shortterm memory and information about intersensory and intrasensory integration which may be useful for obtaining descriptive information about a student in this area. Given the restricted range of behaviors sampled on the test and the limited technical data, the VADS is insufficient for determining a performance level or for making instructional planning and/or evaluation of student progress decisions. Rosenthal (1979) concluded that it adds little to information obtainable from the use of the WISC-R.

VADS should not be used for physically handicapped students who have serious visual, hearing, or motor impairments.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

References

Publisher

Koppitz, 1977
Rosenthal, 1979
Sattler, 1982

Grune & Stratton

SOCIAL/EMOTIONAL ASSESSMENT

Social/emotional assessment is accomplished by a variety of techniques, such as behavioral checklists, projective techniques, self-report measures, situational measures, and observational procedures. Each technique provides information about an individual's social/emotional functioning.

Instruments in this section require extreme caution when interpreting results. Few have included students of different ethnic groups and socio-economic levels in their standardization samples.

Title	<u>Behavior Rating Profile (BRP)</u>	Age	<u>6-1/2 - 14 years</u>
Author	<u>L. Brown & D.D. Hammill</u>	Time	<u>Untimed</u>
Copyright Date	<u>1978</u>	Type	<u>NRT</u>

Intended Purpose of Test

The BRP is designed to rate the behavior of elementary school students across various settings in which the students function.

Description

The BRP contains six components: Student (Self) Rating Scales: Home, School, and Peer; the Teacher Rating Scale; the Parent Rating Scale; and a Sociogram. In the Teacher and Parent Rating Scales, respondents are asked to rate verbal statements which describe student behavior. In the Student Rating Scale, students are asked to classify statements about specific situations about themselves as "True or False". The Sociogram, consisting of eight steps, places the student in relation to his/her classroom peers.

Norm Data

The scales were standardized on an unselected sample of 1,362 students, 645 teachers, and 847 parents in 11 states. Characteristics regarding sex, age, geographic location, residence, occupation, and educational status were described in the manual. No data regarding race were provided.

Reliability

The internal consistency coefficients were reported as ranging from .77 to .98, with 87% of the coefficients exceeding .80. No mention was made of interrater reliability.

Validity

Concurrent validity was based on correlations with the Walker Problem Behavior Identification Checklist, the Quay-Peterson, and the Vineland. Of the 72 coefficients reported, 65% (46 coefficients) exceeded .80 and 86% (62 coefficients) exceeded .35. Diagnostic validity was based on the scale's ability to distinguish between previously identified normal, learning disabled, and emotionally disturbed students.

Special Alerts/Comments

The BRP provides information as to the student's perceived behavioral functioning in school, at home, and with peers. The results of the BRP may be useful for making screening and evaluation of student progress decisions with respect to social behavior.

Alone it is insufficient in determining a student's social/emotional level or functioning and making instructional planning decisions in the social/emotional area. However, it may be used for obtaining supplementary, descriptive information with respect to social/emotional functioning.

Unlike the Devereux, this scale only contains negative statements of behavior. This may bias the rater; also, no behavioral strengths can be tapped. Although the instrument provides an ecological approach to assessing behavior, its usefulness for the New York City population requires local norming. Its use with linguistically and culturally diverse students and LEP students is questionable. Scores should not be reported for bilingual and/or LEP students.

References

Manual
McLoughlin & Lewis, 1981

Publisher

American Guidance Service

Title	<u>Brief Behavior Rating Scale (BBRS)</u>	Grade	<u>K-9th</u>
Author	<u>P. Kahn & S. Rioner</u>	Time	<u>5 minutes</u>
Copyright Date	<u>1978</u>	Type	<u>NRT</u>

Intended Purpose of Test

The purpose of the BBRS is to provide an instrument that can be used for prescriptive teaching recommendations and measuring behavior change over time.

Description

The test consists of 28 items which measure seven personality traits. Each item consists of a question about classroom-observable behavior which may be answered by a special education teacher. Each personality trait is tapped by four questions distributed randomly in the questionnaire.

The personality traits are abrasiveness, inattentiveness, dependency, restlessness, self-involvement, bizarreness, and frustration-tolerance.

Norm Data

The BBRS is a modification of several Devereux scales. Means and standard deviations were presented for students in regular grades and in special education classes. Several hundred students from New York City schools were included in the sample.

Reliability

Internal consistency ranged from .79 to .91 for each subscale and .92 for the total scale.

Validity

The test successfully discriminated among the behavior patterns of students in classes for the neurologically impaired, emotionally handicapped, and socially maladjusted. Diagnostic validity was based on a cross-validation study which indicated that it discriminated between students in regular classes and those in classes for the emotionally handicapped and socially maladjusted.

Special Alerts/Comments

The BBRS provides information about specific classroom-observable behavior with respect to seven personality traits. The results of the BBRS may be useful for making screening decisions with elementary and junior high school students. The findings are insufficient for determining social/emotional functioning, instructional planning, or evaluation of student progress decisions in the social/emotional area. However, as supplementary information, it provides the teacher with a clearer understanding of the student's behavior.

Interpretation with culturally diverse populations and LEP students should be done cautiously. Scores may not be reported for bilingual and/or LEP students.

References

Manual
Kahn & Richner, 1982

Publisher

Paul Kahn

Title Burks Behavior Rating Scales (Burks) Grade 1st -9th
Author M.F. Burks Time Untimed
Copyright Date 1977 Type CRT

Intended Purpose of Test

The Burks is intended to identify a student's behavioral difficulties.

Description

The Burks is a 110-item behavioral rating scale. An adult rater who knows the student well, assesses the student along various behavioral dimensions, rating them on a 5-point scale ranging from "You have not noticed this behavior at all" to "You have noticed the behavior to a very large degree". Nineteen groupings of the 110 items, found by factor analytic studies, are:

Excessive Self-Blame	Poor Impulse Control
Excessive Anxiety	Poor Reality Contact
Excessive Withdrawal	Poor Sense of Identity
Excessive Dependency	Excessive Suffering
Poor Ego Strength	Poor Anger Control
Poor Physical Strength	Excessive Sense of Persecution
Poor Coordination	Excessive Aggressiveness
Poor Intellectuality	Excessive Resistance
Poor Academics	Poor Social Conformity
Poor Attention	

Results of the Burks are recorded on a profile sheet which can be used for diagnostic purposes. Ratings from both parents and the teacher are desirable.

Test Administration

A teacher should observe a student on a daily basis for at least two weeks before rating him or her. When possible parental ratings are recommended to supplement the teacher ratings.

The rating scales are in a four-page booklet. The individual rating the student should not be the person scoring the results. Diagnosis is based on information from the profile sheet.

Technical Data

Content validity was established through scale development procedures; items were selected from clinical observations and from evidence in the literature and were reviewed by a panel of over 200 school psychologists and teachers. Construct validity was demonstrated by significant relationship

between the Burks and the School Attitude Survey. It has also been demonstrated that the Burks has the ability to distinguish between groups of disturbed and normal children.

Reliability was determined through re-rating of 95 emotionally disturbed students in 1st to 6th grade over a 40-day interval. Correlations ranged between .60 and .83; the average item/item retest correlation was .71.

Effective Use /Comments

Although the Burks may be useful for obtaining descriptive information about the students behavioral difficulties, the checklist findings are insufficient for making performance level, evaluation of student progress, and individual planning decisions in the area of behavior.

The Burks is not suitable for screening large groups of students who are performing adequately in the school setting. It is designed to assess students who are already known to be in difficulty.

The manual offers remedial approaches for the nineteen behavioral categories.

Use of the instrument with culturally and linguistically diverse populations and LEP students has not been demonstrated.

Reference

Manual

Publisher

Western Psychological Services

Title	<u>Child Behavior Checklist (CBCL)</u>	Age	<u>4 - 16 years</u>
Author	<u>T.M. Acnenbach & C. Edelbrock</u>	Time	<u>Untimed (15 minutes)</u>
Copyright Date	<u>1983</u>	Type	<u>NAT</u>

Intended Purpose of Test

The CBCL is designed to provide a standardized description of behavioral problems and social competencies among students.

Description

The CBCL consists of a 118-item behavior checklist and a 20 item social competence checklist which is completed by the student, parents, and teachers; there is also a direct observation form.

One checklist is filled out by each parent; adoptive parents, foster parents, relatives with whom the child resides, and others who know the child well and interact as parents do can also fill out the checklist.

Norm Data

Norms were developed from data on 50 children of each sex at each age from 4 to 16. A sample of 1,306 subjects were chosen using census data on socioeconomic, race and age distributions. Urban, suburban, and semirural environment were included in the sample. CBCL ratings were obtained from parents whose children were considered "normal" in the sense that they were not receiving mental health services for at least the preceding year.

Reliability

Intraclass correlations (ICCs) between item scores obtained from mothers filling out the CBCL at 1-week intervals, between mothers and fathers independently filling out the CBCL, and three different interviewers were obtained. All ICC's were .90 or above. The ICC for over a 3-month interval was .84 for behavior problems and .97 for social competence items.

For scale scores and total problems and competence scores, the median Pearson product correlation for 1-week test-retest reliability of mothers' rating was .89.

Validity

The authors designed this instrument to identify children whose behavior problems aroused sufficient concern to warrant referral for professional help. Content validity was demonstrated by the fact that almost all

behavior problem items and all social competence items were significantly associated with clinical studies.

Correlations of scores with other recognized behavior checklists resulted in coefficients in the middle and highest magnitudes of r , indicating that the CBCL is measuring the same constructs. Discriminant analysis and other regression procedures for studying score differences between clinic and nonreferred populations also contributed evidence for the construct validity of the scales.

Special Alerts/Comments

The CBCL provides a standardized description of a student's behavioral problems and social competencies which is useful for obtaining supplementary descriptive information about social-emotional functioning. With other measures, it may be used to make screening, performance level, evaluation of student progress, and instructional planning decisions developing short- and longterm goals on the IEP.

The CBCL is well standardized and has adequate reliability and validity. In fact, some experts in the field feel that the CBCL is an exemplary test, likely being the best instrument of its kind in use today. The test is inexpensive and easy to administer and score. Where a behavioral checklist is needed, the CBCL should be the instrument of choice.

No scale of the CBCL is directly equivalent to any clinical diagnosis. Diagnostic labels should not be assigned on the basis of a high score on any of the CBCL's behavior problem scales.

Results with linguistically and culturally diverse and LEP students should be interpreted cautiously.

References

Manual
Mitchell, Jr., 1985

Publisher

Thomas M. Achenbach

Title	<u>Children's Human Figure Drawings (HFD)</u>	Age	<u>5 years to Adult</u>
Author	<u>E. Koppitz</u>	Time	<u>Untimed</u>
Copyright Date	<u>1968</u>	Type	<u>Projective</u>

Intended Purpose of Test

The HFD is designed to assess how a student feels about him/herself and to obtain an index of the student's "mental maturity."

Description

The student is asked by the examiner to "draw a whole person." No age or sex is specified. Drawings should be made on a blank sheet of 8-1/2" X 11" paper with a number two pencil with an eraser. Koppitz's scoring system, presented in her book Psychological Evaluation of Children's Human Figure Drawings, contains criteria for evaluating a drawing for its Developmental Indicators (related to age and maturation) and Emotional Indicators.

Norm Data

Norms were based on HFD's of 1,856 public school students, kindergarten through 6th grade, attending 86 classes in ten elementary schools located in a mid-western city, three small mid-western and eastern towns and in rural areas in the mid-west and east. One-third were from low socioeconomic black and white families, one-third from white middle income families and one-third from high income families. Data were analyzed by age and sex.

Reliability

Reliability of scoring human figure drawings for developmental items and for emotional indicators was established by independent judges scoring the same protocols; 467 items were checked and 95% of them received unanimous agreement. No other type of reliability was mentioned.

Validity

Research studies conducted by Koppitz have shown that human figure drawings differentiate significantly between children with certain kinds of emotional problems and children in whom no such handicapping conditions exist. Differences in drawings between clinic and non-clinic populations were also found. Comparisons between HFD and the S-B and WISC ranged from .45 to .80 depending upon age. Studies were also conducted using HFD to help identify brain-injured students.

Special Alerts/Comments

The HFD provides an indication of how a student feels about him/herself. Because of the technical limitations of projective tests, the HFD does not provide sufficient data for determining a level of performance or for making instructional planning or evaluation of student progress decisions. As an informal tool, it may provide descriptive information to clarify the clinical picture of the personality where a disturbance is strongly suspected (Buros, 1972). The HFD should only be administered and scored by a school psychologist. The test should not be used with students having serious motor difficulties.

Its use with linguistically and culturally diverse students and LEP students is questionable. Scores may not be reported for bilingual and/or LEP students.

References

Buros, 1972
Swenson, 1965
Koppitz, 1968

Publisher

Grune & Stratton

Title	<u>Conners Rating Scales (CRS)</u>	Grade	<u>1st - 8th</u>
Author	<u>C.K. Conners</u>	Time	<u>Untimed</u>
Copyright Date	<u>1973, 1982</u>	Type	<u>NRT</u>

Intended Purpose of Test

The CRS is intended to help identify hyperactive students and evaluate treatment effectiveness.

Description

The CRS contains Parent and Teacher Rating Scales. Two versions of each are available. The original Conner scale (1973) contains a 93-item Parent and a 39-item Teacher Scale. The Revised (1982) edition contains a 48-item Parent and a 28-item Teacher Rating Scale. The author recommended continued use of the longer scale, but conceded that clinicians may prefer the shorter version since no sex-by-age normative data exist for the original scale.

Eight factors have been identified through factor analysis of the 93-item scale. These are: conduct disorder, fearful-anxious, restless-disorganized, learning problem-immature, psychosomatic, obsessional, antisocial, hyperactive-immature. Five primary factors are: found on the revised Parent Rating Scale, these were: conduct problem, learning problem, psychosomatic, impulsive-hyperactive, anxiety. The 28 item Teacher Rating Scale has three primary factors: conduct problem, hyperactivity, inattentive-passive.

Both original and revised version contain a "Hyperactivity Index".

Norm Data

A sample of 683 children between the ages of 6 and 14 participated in the standardization of the full scale. Sex-by-age normative data are not available for the full scale version of the CRS.

Normative data for the revised versions of the Parent and Teacher Rating Scales were based on a sample of 570 children, ages 3 to 17.

Reliability

Conners (1973) reported a 1- month test-retest reliability ranging from .70 to .90 for the 39 item Teacher Rating Scale. Over a 1- year test-retest interval, reliabilities ranged from .30 to .70. Interrater correlations were highly significant ($p < .001$) with no significant differences between mother and father ratings. No reliability data were reported for the revised edition of the CRS.

Validity

Numerous studies have demonstrated that the CRS discriminates between normal and hyperactive children.

Special Alerts/Comments

The CRS are used to identify hyperactive students and should not be used as the sole criterion for establishing such a diagnosis. With other measures, findings may be useful for determining social emotional functioning, instructional planning, and evaluation of student progress.

The information provided by the CRS is sufficient for developing shortterm objectives on the IEP in the area of social/emotional. Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

Reference

Manual

Publisher

C.K. Connors

Title	<u>The Devereux Elementary School Behavior Rating Scale (DESB)</u>	Grade	<u>K-6th</u>
Author	<u>G. Spivack & M. Swift</u>	Time	<u>10 minutes</u>
Copyright Date	<u>1967</u>	Type	<u>NRT</u>

Intended Purpose of Test

The scale is designed for use by classroom teachers to describe and understand the behavior problems of their students which interfere with successful academic achievement.

Description

The DESB consists of 11 subscales: Classroom Disturbance, Impatience, Disrespect-Defiance, External Blame, Achievement Anxiety, External Reliance, Comprehension, Inattentive-Withdrawn, Irrelevant Responsiveness, Creative Initiative and Need for Closeness to the Teacher, each of which contains four to five behavioral items. The teacher rates each item with respect to its observed frequency.

Norm Data

The sample consisted of 252 children in residential treatment facilities, 100 mentally retarded children, and 348 public school students.

Reliability

Authors reported test-retest reliability based on a 1-week interval as .87. Schaefer, et al (1975) reported interrater reliability as .87.

Validity

The items of the scale indicate the content validity of the DESB. No other types of validity were reported.

Special Alerts/Comments

The results of the DESB may be useful for making screening decisions with respect to social behavior and to obtain supplementary, descriptive information about a student's classroom behavior. Alone, given the nature of the sample, test findings are insufficient for determining social/emotional functioning, instructional planning, and evaluation of student progress decisions in the social/emotional area.

Because of the limited normative sample, its use with linguistically and culturally diverse students and LEP students is questionable. Scores may not be reported for bilingual and/or LEP students.

References

Manual
Schaefer, Baker,
& Zavel, 1975

Publisher

Devereux Foundation Press

Title	<u>Piers-Harris Children's Self Concept Scale (CSCS)</u>	Grade	<u>3rd - 12th</u>
Author	<u>E. Piers & D. Harris</u>	Time	<u>15 - 20 minutes</u>
Copyright Date	<u>1969</u>	Type	<u>NRT</u>

Intended Purpose or Test

The CSCS was developed to assess the attitudes and feelings a student holds about him/herself.

Description

The CSCS consists of 80 first-person statements (e.g., "I am a happy person", "I am a sad person"). Half of the items reflect positive self concept and half reflect negative self concept. Students respond to each statement by indicating yes or no. For young students, the items are read aloud by the examiner. The CSCS may be administered to a group or individually.

Norm Data

The scale was standardized on 1,183 boys and girls in grades 4 to 12 in one Pennsylvanian school district. The sample reflects a cross-section of the socioeconomic status levels of the district. Information regarding the racial breakdown was not cited in the manual.

Reliability

Internal consistency coefficients (KR21) ranged from .78 to .93 for the various grades. Test-retest (at a 4-month interval) reliability coefficients ranged from .71 to .77.

Validity

The CSCS's concurrent validity was based on correlations with similar instruments (e.g., Lepsett's Children's Self Concept Scale, .64) and with teacher and peer ratings (.40).

Special Alerts/Comments

The CSCS provides an indication of a student's attitudes and feelings about him/herself. Although designed as a research tool, the results of the CSCS may be useful for obtaining supplementary, descriptive information about a student's self concept.

Given the technical inadequacies of the CSCS, the test findings are insufficient for determining a student's social/emotional functioning, instructional planning, and evaluation of the student progress decisions in social/emotional area.

Since the description of the standardization sample was limited, use of the CSCS with handicapped and culturally and linguistically diverse students and LEP students is questionable. Scores may not be reported for bilingual and/or LEP students.

References

Manual
Buros, 1978

Publisher

Western Psychological Services

Title	<u>Revised Behavior Problem Checklist (RBPC)</u>	Age	<u>5 - Adolescent</u>
Author	<u>H.C. Quay & D.R. Peterson</u>	Time	<u>Untimed</u>
Copyright Date	<u>1983</u>	Type	<u>MRT</u>

Intended Purpose of Test

The RBPC is designed to identify problem behaviors commonly seen in children and adolescents.

Description

The RBPC is an 89-item questionnaire requiring the respondent to rate each item as either not a problem, a mild problem, or a severe problem. The test is self-administered and requires that the respondent be someone who knows the student well (e.g., parent, teacher, child care worker). Scoring takes approximately five minutes with the use of a scoring stencil.

The test consists of the following six subscales: conduct disorder (CD), socialized aggression (SA), attention problems-immaturity (AP), anxiety-withdrawal (AW), psychotic behavior (PB), and motor excess (ME).

Norm Data

RBPC users are encouraged to develop their own local norms. The authors, however, did provide some data which can serve as reference points for the analysis of subscale scores. However, these data may be of limited usefulness because of the composition of the sample. Inner city minority students and special education students were not represented. The sample consisted of more than 90% white students of predominately working and middle-class families.

Reliability

Measures of internal consistency (alpha reliabilities) ranged from .70 to .95. In a sample of 149 first through sixth grade students, tested after a 2-month interval, subscale test-retest reliabilities ranged from .49 to .83; in this study only the AP scale had adequate reliability.

Validity

Concurrent validity was determined by contrasting scores of normal children (293 males and 273 females) with clinical cases (66 males and 33 females). The differences between the means were substantial for both sexes. All differences, except one, were significant beyond the .01 level.

Special Alert/Comments

The RbPC provides information about problem behaviors which may be, useful for screening decisions. Given the limited standardization sample and weak reliability data, results may not be used for making performance level, instructional planning, and evaluation or student progress decisions. The results may provide useful information for developing shortterm objectives on the IEP. The RbPC can also be used for assisting with clinical diagnoses and classifying juvenile offenders. Local norms should be developed.

Interpretation of scores with culturally and linguistically diverse populations and LEP students should be done cautiously. Scores may not be reported for bilingual and/or LEP students.

References

Manual
Mitchell, Jr., 1985

Publisher

Herbert Quay

Title	<u>Rorschach Psychodiagnostic</u>	Age	<u>3 years to Adult</u>
Author	<u>M. Rorschach</u>	Time	<u>20-60 minutes</u>
Copyright Date	<u>1921</u>	Type	<u>Projective</u>

Intended Purpose of Test

The Rorschach is designed to assess perceptual-cognitive style in which the individual's problem-solving approach reflects key dimensions of his/her traits. Trained psychologists appraise both intellectual and social/emotional aspects of personality functioning in adults and children. Based on a psychodynamic model, it is intended to identify one's intrapsychic conflicts, as well as the individual's approach to resolving these conflicts.

Description

The Rorschach consists of ten ambiguous inkblots, printed on cards, five of which are achromatic while the other half include color. Administration involves presenting each card to the student who is then requested to tell what he/she sees or what the inkblot looks like (free association). Responses are recorded verbatim. The student is then presented with each card again and asked to indicate what parts of the inkblot he/she used in arriving at a response (inquiry). There are a variety of highly sophisticated scoring systems available for interpreting responses (see (Exner, 1978) which are not included with the test materials.

Norm Data

The Rorschach has been normed on several age groups and types of populations since its development in 1921. Norms for various types of responses given by a sample of 4,000 children ranging in age from five to sixteen years old were discussed by Exner (1978; 1986). The sample was not controlled for SES, conduct and personality disorders.

Reliability

Test-retest reliability was reported in numerous studies (Buros, 1972, 1978) and ranged from .15 to .93. Interscorer reliability between specially trained clinicians was generally greater than .90 (Exner, 1986).

Validity

Validation of the Rorschach was generally based on psychodynamic theory and showed that, as a diagnostic tool, it can differentiate between students exhibiting different types of problems and functional levels (e.g., introversion-extroversion; high and low M.A.) (Exner, 1986). Concurrent validity coefficients based on other measures of personality ranged from .55 to .83 (Ames, 1952).

Special Alerts/Comments

The Rorschach provides an indication of intellectual and social/emotional aspects of personality functioning. Because of the technical limitations of projective tests and the variety of scoring techniques available to trained examiners, (Weiner, 1986), the Rorschach does not provide sufficient data for determining a level of performance or for making instructional planning or evaluation of student progress decisions. As an informal tool, it may provide information with respect to the development of insight and understanding of the personality dynamics of the student as a result of the sample of patterned interactions between the perceptual, cognitive, and emotional components. The Rorschach may only be used, administered, and interpreted by a school psychologist with specific training in projective tests. It is more appropriately used with verbal students, but may be quite useful when assessing withdrawn students. In all instances, it should be used with caution.

Although recent research (Scott, 1981) suggested that the Rorschach holds promise as a diagnostic tool which minimizes test bias when assessing students from culturally and linguistically different backgrounds, its use with such students and LEP students has not been demonstrated.

References

Buros, 1972, 1978
Ames, 1952
Exner, 1978, 1986
Scott, 1981
Weiner, 1986

Publisher

The Psychological Corp.

Title	<u>Social-Emotional Dimension Scale</u>	Grade	<u>K - High School</u>
	<u>(SEDS)</u>		
Author	<u>J.B. Hutton & T.G. Roberts</u>	Time	<u>Observation Period of</u> <u>three weeks</u>
Copyright Date	<u>1986</u>	Type	<u>NRT</u>

Intended Purpose of Test

The SEDS measures nonacademic behaviors which may be interfering with a student's own learning and the learning of others.

Description

The SEDS examines school difficulties from a cognitive/behavioral point of view. The scale is based on four fundamental assumptions: Behavior is learned; behavior is a function of the setting within which it occurs; behavior is a consequence of thought and feelings; and interpretation of behavior is a function of the perception of the observer.

The SEDS is a 32-item rating scale divided into six sections: avoidance of peer interaction, aggressive interaction, avoidance of teacher interaction, inappropriate behavior, depressive reaction, physical/fear reaction. The rater indicates whether the behavior occurs rarely or never, occasionally, or frequently. Independent ratings by different personnel can help ascertain how a student's behavior is perceived in different school settings.

The SEDS yields raw scores, percentiles for each behavior dimension (BQ), a composite percentile, and a social/emotional dimension quotient. A student's social/emotional dimension quotient can be used to determine if the students may have behavioral problems which interfere with the student's own educational program or the education program of others. BQs range from levels indicative of students at very high risk for problematic behavior to average levels of risk. The quotient scores serve to emphasize the presence of risk behavior.

Norm Data

The SEDS was standardized on a sample of 1,097 students ranging in age from 5 1/2 to 19 years by 198 raters across a geographic cross-section of the country. The dimension of sex, urban/rural, race, ethnicity, geographic distribution, and educational levels were reflective of percentages of the U.S. population as reported in the 1985 Statistical Abstract.

Reliability

Average internal consistency was reported as .88.

Validity

The validity of SEDS was established with respect to content, criterion related, and construct validity. Content validity was built into the scale by reviewing definitions pertaining to emotional disturbance and behavior disorders. Criterion related validity was established by high correlation coefficients with the Walker. Construct validity was established by demonstrating that the SEDS differentiated between normally achieving and mildly handicapped subgroups of mildly mentally retarded, learning disabled and emotionally disturbed students.

Special Alerts/Comments

The SEDS provides information about problem behaviors which may be useful for screening decisions. Given the limited technical data, results may not be used for making performance level, instructional planning, and evaluation of student progress decisions.

The SEDS may be useful for obtaining descriptive information about the student's social/emotional functioning and for developing shortterm objectives on the IEP in the area of social/emotional development. The manual provides case study examples and intervention suggestions.

Use of the SEDS with culturally and linguistically diverse populations and LEP students has not been demonstrated.

Reference

Manual

Publisher

PPC-ED

Title	<u>STS Youth Inventory (STS)</u>	Grade	<u>7th - 12th</u>
Authors	<u>H.M. Remmers & B. Shimoerg</u>	Time	<u>Untimed (30 - 45 minutes)</u>
Copyright Date	<u>1971</u>	Type	<u>NRT</u>

Intended Purpose or Test

The STS is designed to identify a student's needs and problems as he/she sees them.

Description

The STS comes in booklet form and is accompanied by a manual which provides administrative, interpretive, and technical information. A Spanish edition of the inventory is also available.

The STS is organized into five categories.

<u>Category</u>	<u>Contents</u>
1. My School	problems that may keep a student from doing his/her best.
2. After High School	problems of self-appraisal and career planning.
3. About Myself	self-concept and problems student has or thinks he/she has with personal adjustment or happiness.
4. Getting Along with Others	relationships with classmates and adults.
5. Things In General	Attitudes and values related to social and moral problems.

The items are written in simple, everyday language. The student is required to indicate the intensity of the problem by marking a response in a box or circle. The inventory may be administered individually or as a group.

Norm Data

Approximately 4,000 students from 4-year high schools in 12 states participated in Form G, the present revision. No information on sample characteristics was included.

Reliability

Split-half reliability coefficients ranged from .86 to .96 for individual categories.

Validity

Content validity was demonstrated through the development of the inventory. Construct validity was developed through factor analysis in which four factors with r 's .50 or above were identified.

Special Alerts/Comments

The STS provides information about a student's needs and problems which may be useful for obtaining a descriptive report about the student's social/emotional functioning. Given the limited technical data, the findings are insufficient for making screening, performance level, evaluation of student progress, and instructional planning decisions in this area. With other measures, results may be used to develop shortterm objectives on the IEP for social/emotional functioning.

Use with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

Reference

Manual

Publisher

Scholastic Testing Service

Title	<u>Thematic Apperception Test (TAT)</u>	Age	<u>7 years to Adult</u>
Author	<u>H. Murray</u>	Time	<u>Untimed</u>
Copyright Date	<u>1943</u>	Type	<u>Projective</u>

Intended Purpose of Test

The TAT is a technique designed for investigating "the dynamics of personality as it manifests itself in interpersonal relations and in the apperception or meaningful interpretation of the environment" (Beilak, 1975).

Description

The TAT is an individually administered measure consisting of a series of 31 pictures of people engaged in a variety of activities. The person taking the test is asked to tell a story about each picture, indicating what is happening in the scene, what led up to it, what the people are thinking and feeling, and what the outcome of the story will be. All responses are recorded verbatim. Although a number of different scoring systems have been developed, in actual practice, the test is seldom scored. Rather, the clinician determines major themes, conflicts, and dynamics present in the stories.

Norm Data

Although the manual presented no data regarding normative standardization, subsequent research studies have provided normative data on several hundred samples (Buros, 1978).

Reliability

The manual provided no information regarding reliability. However, several studies have reported split-half, test-retest, and interrater reliabilities; results varied widely.

Validity

Although the manual did not report on the validity of the TAT, a large body of research has focused on the relationship between performance on the TAT and real life behavior (Buros, 1978). The results have been equivocal.

Special Alerts/Comments

The TAT provides an indication of social/emotional behavior. Because of the technical limitations of projective tests, the TAT does not provide sufficient data for determining a level of performance or for making instructional planning or evaluation of student progress decisions. As an informal tool, it may provide descriptive information with respect to

social/emotional development. The TAT can only be administered and interpreted by a school psychologist with specialized training in projective testing. The TAT appears useful in providing insights about withdrawn and aggressive students.

Results with linguistically and culturally diverse students and LEP students should be interpreted with caution, since no extensive research has been reported using the TAT with these students.

References

Publisher

Manual
Buros, 1978
Beilak, 1975
Anastasi, 1976

Harvard University Press

Title	<u>walker Problem Behavior Checklist (Walker)</u>	Grade	<u>Preschool - 6th</u>
Author	<u>H.M. Walker</u>	Time	<u>Untimed</u>
Copyright Date	<u>1983</u>	Type	<u>NRT</u>

Intended Purpose of Test

The Walker is designed to identify students having behavioral difficulties who should be referred for evaluation, and treatment.

Description

The walker is a 50-item checklist which measures problem behaviors in five areas: Acting Out, Withdrawal, Distractibility, Disturbed Peer Relations, and Immaturity. The checklist is completed by teachers following a 2-month observation period. The teacher records whether a checklist item has or has not occurred in the 2-month interval.

Each item carries a weight representing the handicapping influence of that behavior on the student's adjustment to school. Score weights are summed across the five problem areas and then converted to T scores. A cut-off score of 60 or higher suggests a need for further evaluation.

Norm Data

The Walker was normed on three samples: preschool/kindergarten (N=469) grades 1 through 3 (N=852), and grades 4 through 6 (N=534). Sample sizes are adequate; however, sampling procedures and the characteristics of the norm groups were not specified in the manual. Geographic diversity of the sample was inadequate.

Reliability

Split-half reliability was reported as .98. Test-retest coefficients ranged from .43 to .88. Interrater reliability was not reported.

Validity

Content validity was based on descriptions of problem behavior which were developed by teachers. Criterion validity was demonstrated from studies in which walker scores for behaviorally disturbed students were significantly higher than score of non-behaviorally disturbed students. Construct validity was also demonstrated through factor analysis.

Special Alerts/Comments

The Walker provides information about behavioral difficulties which are useful for making screening decisions. Given the limited technical data, the information is not sufficient for making performance level, instructional planning, and evaluation of students progress decisions. With other measures, it may be used for developing shortterm objectives on the IEP for social/emotional functioning.

A weakness of the Walker is that it fails to evaluate the frequency, and duration, and latency and interresponse time of the behaviors. These important behavioral dimensions would provide a more comprehensive description of behavioral difficulties.

Use with linguistically and culturally diverse students and LEP students has not been demonstrated.

Reference

Manual
Mitchell, Jr., 1985

Puolisher

Western Psychological Services

TESTS FOR BILINGUAL AND/OR LIMITED ENGLISH PROFICIENCY (LEP) STUDENTS

Tests for language dominance, English language proficiency, and native language proficiency are examined in this section. In estimating which language is stronger, language dominance instruments provide a comparison of the proficiency in the languages spoken by the student and his/her degree of bilingualism. Note that determination of Limited English Proficiency (LEP) status is based on the Language Assessment Battery (LAB) for all students in New York City.

Cognitive and achievement tests in languages other than English are also examined in this section. The examiner must be fluent in the languages being tested. Due to technical inadequacies of standardization procedures, most tests are insufficient for reporting scores with linguistically and culturally diverse students.

Note that the results of tests other than the LAB may not be used to determine if the student should be assessed by a bilingual or a monolingual team. The LAB is the only instrument used for this purpose in New York City.

In addition to using this section as a guide for selecting tests for assessment of bilingual and LEP students, clinicians should review tests in other sections of this manual.

Title	<u>Basic Inventory of Natural Language (BINL)</u>	Age	<u>K - 12th</u>
Author	<u>C. Herbert</u>	Time	<u>Untimed (approximately 30 - 40 minutes)</u>
Copyright Date	<u>1973</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

The BINL is a language system designed to assess the language dominance, proficiency, and level of complexity in English, Spanish, and 17 other languages.

Description

The BINL consists of 40 large story starter pictures with a sequence book, talk titles, profile sheets, and scoring forms. A standard tape recorder is used to tape student responses. The test is administered individually. The examiner tries to elicit natural, spontaneous, verbal responses using three to five pictures as stimuli, which may be selected by the individual student. The scoring is based upon fluency (number of words in each sentence, up to a 10-sentence response), and complexity (use of different parts of speech, completeness of sentences, and use of phrases and subordinate clauses).

Norm Data

Approximately 400 English-dominant and Spanish-dominant students, in lower elementary grades, who lived in Southern California and Illinois, comprised the standardization sample. Scores were reported for fluency and level of complexity.

Reliability

Reliability on stability and accuracy of the language system was not reported. However, one manual cited two studies which found that students dominant in either English or Spanish do generate highly consistent levels in oral language complexity across the ten sentences spoken on the test ($r = .93$ for each sample).

Validity

Three validity studies were conducted on the BINL in 1974 on English-dominant and Spanish-dominant students from Southern California. Significant correlation coefficients ranged from .26 to .43.

Special Alerts/Comments

The BINL provides an indication of oral language dominance and may be used to obtain descriptive information with respect to a student's expressive language. Given the technical limitations and intended purpose of the test, the results are insufficient for determining language dominance or proficiency.

The administrator must be proficient in the languages in which the instrument is given. To be useful in New York City, local norms need to be established. Caution should be taken when interpreting test scores, since performance may be influenced by dialectal variations.

The results of this test may not be used to determine if the student should be assessed by a bilingual or a monolingual team. In New York City, the Language Achievement Battery (LAB) is the only instrument used for this purpose.

References

Manual
Pletcher, et al, 1978
Erickson & Omark, 1981

Publisher

CHECpoint Systems, Inc.

Title Bateria Woodcock Psico-Educativa en
Espanol (Bateria)

Age 4 to 19 years

Author R. W. Woodcock

Time Untimed
(5 minutes per subtest)

Copyright Date 1982

Type NRT

Intended Purpose of Test

The Bateria is designed to measure cognitive functions and academic achievement of Spanish-speaking students.

Description

The Bateria consists of 17 subtests organized into two sections: cognitive ability and achievement. It is individually administered in its entirety or for selected subtests. The complete Bateria includes a 444-page flip-book in an easel format, Examiner's Manual, and 16-page response booklet. Two subtests require the use of a watch with a second hand.

A parallel form of the Bateria in English, the Woodcock-Johnson Psychoeducational Battery (W-J Battery), is available.

Norm Data

Standardization was based on 802 students from kindergarten and grades 1, 3, 5, 8, and 11 from urban areas in Costa Rica, Mexico, Peru, Puerto Rico, and Spain. The sample, randomly chosen from public and private schools in 1980, participated in all 17 subtests.

Results are reported as age/grade equivalents, percentile ranks, standard scores, functional levels, instructional ranges, normal curve equivalents, and relative performance indexes. Subtest scores may be grouped into clusters. Achievement-aptitude differences are also reported. Age norms are provided for ages 4 to 19, and grade norms are provided for kindergarten through 12.9.

Reliability

Split-half reliabilities corrected by the Spearman-Brown formula were calculated for all subtests except "spatial relations" and "visual matching", which were based on studies done on the W-J Battery (Woodcock, 1978, p. 119). Over 100 subtest coefficients, reported at varying grade levels, ranged from .55 to .98.

The reliability of clusters calculated by Mosier's formula ranged from .77 to .92. Full scale reliabilities ranged from .94 to .96.

Validity

Content validity was established through test construction and development. Construct validity was indicated by the intercorrelation patterns of the subtests. Criterion-related validity was indicated by comparing the Bateria with the Mexican version of the WISC-R; Fletcher (doctoral dissertation in preparation) reported correlation coefficients of .86 and .84 for two sample groups.

Special Alerts/Comments

The Bateria is designed to measure cognitive functions and academic achievement of Spanish-speaking students. The information provided on cognitive functions is insufficient for making performance level decisions in this area. The information provided on achievement is sufficient for making evaluation of student progress and instructional planning decisions, and developing short- and longterm objectives on the IEP in the areas of reading and math. Norms must be used cautiously because of the standardization sample.

Scores should not be reported for LEP and/or bilingual students.

References

Manual
Technical Summary
Woodcock, 1978
Fletcher (In Press)

Publisher

DLM Teaching Resources

Title	<u>Bilingual Syntax Measure I (BSM)</u>	Age	<u>4-9 years</u>
Author	<u>K. Burt, H. Dulay, & E. Hernandez</u>	Time	<u>Untimed</u> <u>(15-20 minutes)</u>
Copyright Date	<u>1975</u>	Type	<u>NRT</u>

Intended Purpose of Test

The BSM is intended to determine the student's comparative language dominance or oral proficiency in Spanish and English.

Description

The BSM consists of manuals (English and Spanish), a technical handbook, and color plates. Students respond orally to questions asked about the drawings, and responses are recorded. The responses are then analyzed according to the BSM standards of grammar and acceptability. English and Spanish versions may be administered separately, and are parallel tests rather than translations. Scores yield classification in one of five hierarchical categories of proficiency.

Norm Data

The standardization employed over 1,000 students for each version (English and Spanish) of the tests; all were enrolled in grades kindergarten to 2 in four geographic regions of the U.S. The data were considered illustrative rather than normative.

Reliability

Based on a sample of 150 Spanish-speaking students, test-retest reliability was determined using the Kappa coefficient (.4). Standard errors of measurement tended to be one or less. Interscorer reliability was higher than test-retest. Lower reliability of the Spanish version may be related to disagreements regarding dialectal variation.

Validity

Construct validity was supported by a body of psycholinguistic research described in the Technical Manual outlining the hierarchical nature of the acquisition of specific syntactical structures sampled by the BSM. The scoring system employed by the BSM is based on the available literature on this hierarchy and on the variability reported in research studies. Results of factor analysis for the various structures in both languages are presented in the Technical Handbook.

Special Alerts/Comments

The BSM provides an indication of language dominance and oral proficiency in Spanish and English. The results may yield supplementary, descriptive information with respect to a student's oral proficiency in Spanish and English.

Given its technical limitations, the results of the BSM are insufficient for ~~ost~~termining language dominance or the student's level of language proficiency in Spanish or English.

Scores should not be reported for LEP and/or bilingual students.

The administrator must be fluent in Spanish and English, and familiar with dialectal variations in both languages.

The results of this test may not be used to determine if the student should be assessed by a bilingual or a monolingual team. In New York City, the Language Achievement Battery (LAB) is the only instrument used for this purpose.

References

Manual
Pietchner, et al, 1978
Erickson & Onark, 1981

Publisher

The Psychological Corp.

Title Comprehensive Tests of Basic Skills, Spanish Edition Grade 1st - 8th
(CTBS Español)

Author Norwalk - La Girada Unified School District, Time 2 to 2 1/2 hours (per level)
Southern California

Copyright Date 1978 Type NRT

Intended Purpose of Test

CTBS Español is designed to estimate basic skills achievement in reading and math for Spanish-speaking students.

Description

CTBS Español is a Spanish language adaptation of the CTBS/S Reading and Mathematics tests. The adapted items were field tested and refined before the standardized edition was developed. Although the CTBS Español is an adaptation rather than a translation, it was designed to keep intact the contents and measurement features of the CTBS/S.

CTBS Español consists of five levels, B through 3, corresponding to grades 1 through 8. The test evaluates the following skill areas: Word Recognition, Reading Vocabulary, Reading Comprehension, Mathematics Computation, Mathematics Concepts and Applications. Spanish/English Reading and Vocabulary Screening (SERVS), an optional pretest-practice test, is available separately and is recommended for grades 1 and 2. The SERVS helps to determine whether a student should be tested with the Spanish or English version of the CTBS test, and gives bilingual practice in the mechanics of test-taking.

Norm Data

CTBS Español reading and math tests were standardized on 3,173 and 2,328 students, respectively, in grades 1 to 8 who were biliterate in Spanish. Biliteracy was determined according to a set of criteria which included performance on the SERVS and teacher evaluations. The students were selected from 326 school districts with bilingual programs, receiving ESEA Title VII funds, across the continental United States.

The scores on both language versions were equated and then compared to the national norms on the CTBS/S. "The percentiles are not norms for CTBS Español; rather ... they are estimates of the percentile a student with a given score on CTBS Español would have received if CTBS/S had been taken" (p. 33). Since CTBS Español was normed in the Spring the most accurate performance estimates will be obtained through Spring administration.

Test results are reported as percentiles, stanines and grade equivalents.

Reliability

KR 20 reliability coefficients for individual skill areas across all grade levels were above .80. No other reliability data were provided.

Validity

Little specific information about validity was provided, although there is the implication that the battery has content validity since it follows the content of the original English battery. A study indicated that students earned almost identical math computation test scores on CTBS Español and CTBS/S. However, no content validity study has been conducted on CTBS Español in reading; since the levels of difficulty and hierarchy of skills differ between English and Spanish reading, the assumed content validity is questionable.

Special Alerts/Comments

CTBS Español is designed to estimate basic skills achievement in reading and math for Spanish speaking students which may provide supplementary information on skills in Spanish. Given its limited technical information, CTBS Español results are insufficient for making screening, performance level, instructional planning, and evaluation of student progress decisions.

Although the publisher assumed that performance on CTBS Español predicts performance in English if the testees were biliterate, this assumption cannot be validly sustained for individual students. The only valid conclusion regarding a percentile score on CTBS Español is that the individual student scores above or below the median, or that he or she performs like "x" percent of the students on whom the battery was normed.

Caution should be exercised when interpreting information resulting from this test, particularly in the area of inferential reading. A study by Cabello (1984) suggests that some of the items may be culturally biased as a result of keeping the content of the English test intact when translating/adapting it into Spanish. To assume equality of experiences with a particular topic across different cultures is an erroneous premise leading to bias in testing.

The information provided by the CTBS Español is sufficient for developing shortterm objectives on the IEP in the areas of reading and math. Further diagnostic information may be obtained, but scores should not be used to make placement decisions.

Scores may not be reported for LEP and/or bilingual students.

References

Manual
Technical Report
Cabello, 1984

Publisher

CTB/McGraw Hill

Title Crane Oral Dominance Test

Grade Pre-K to 3rd

Author B.J. Crane

Time Untimed (20 minutes)

Copyright 1976

Type NRT

Intended Purpose of Test

The purpose of the Crane is to determine oral language dominance. It was designed to be culturally and linguistically appropriate for Cuban, Mexican-American, and Puerto Rican students in grades kindergarten to 3.

Description

The Crane is individually administered and is comprised of eight Spanish-English Word Pair Sets. Each set contains four word pairs in English and four in Spanish which are listed randomly. The examiner reads words out loud; the student must recall them in either language. Also, the student is asked five questions referring to home language usage.

Norm Data

In 1975, the test was piloted with 934 Hispanic and Anglo-American students, four to eight years old, living in Arizona, California, New York, and Washington.

Reliability

An odd-even reliability coefficient of .86 was reported.

Validity

A predictive validity coefficient of .87 was reported without identifying the criterion measure used.

Special Alerts/Comments

The Crane provides an indication of oral language dominance. The results may be used to obtain supplementary, descriptive information with respect to a student's oral language dominance in Spanish and English. Given its technical limitations, the results are insufficient for determining the student's dominant language. It may be appropriate as a criterion referenced tool. The complexity of the task may be too taxing for pre-kindergarten students.

The results of this test may not be used to determine if the student should be assessed by a bilingual or a monolingual team. In New York City, the Language Assessment Battery (LAB) is the only instrument used for this purpose.

References

Manual
Los Angeles County, 1978

Publisher

National Educational Laboratory
Publishers, Inc.

Title	<u>Diagnostic Assessment of Basic Skills - Spanish Edition (Brigance in Spanish)</u>	Grade	<u>Pre K - 6th</u>
Author	<u>A.H. Brigance</u>	Time	<u>Untimed</u>
Copyright Date	<u>1983</u>	Type	<u>CRT</u>

Intended Purpose of Test

The Brigance in Spanish is a diagnostic system designed to assess English and Spanish language proficiency and reading and math skills in Spanish.

Description

The Brigance in Spanish is an individual or group administered test consisting of 102 subtests contained within 10 areas: Readiness, Speech, Functional Word Recognition, Oral Reading, Reading Comprehension, Word Analysis, Listening, Writing and Alphabetizing, Numbers and Computation, and Measurement. The manual recommends administering tests at the student's corresponding grade level in order to identify quickly the student's approximate skill levels and the level for subsequent assessments. Supplementary assessments on using the calendar, English and metric linear measurement, and the denominations of coin and paper money are made available but are not used for determining performance levels.

Directions to the students are in Spanish and, for the language dominance assessments, in English as well. The directions for the examiner are in English. Reproducible student record forms are available by grade, kindergarten through six.

Test Administration

The manual indicates criteria for determining areas and tests to administer according to the purposes of the assessment. Proficiency in the basic skills is to be determined by administering any of the 102 tests, as required. Although the manual states that adaptations of the assessment procedures can be made in order to give the student every opportunity to master the tasks, it does not specify which adaptations may be considered; it stipulates, however, that coaching should not be done.

Technical Information

The instrument is referenced to "the most widely used" Spanish and English basal programs nationwide "to validate the scope, sequence, and grade levels of these skills" (p. 187). Each assessment has been reviewed for content validity by a national group of advisors, and has also been field-tested with Spanish speaking students by over 150 professionals in the fields of bilingual, ESL, migrant, and bilingual special education.

No psychometric or additional technical data are provided.

Effective Use/ Comments

The Brigance is a diagnostic system designed to assess English and Spanish language proficiency and reading and math skills in Spanish which, with other measures, may be useful for determining performance level and instructional planning decisions including the development of short- and longterm objectives on the IEP.

Skill assessments, such as for word analyses and letter-sound correspondence, are not related to grade level "because there is no consensus among the Spanish texts referenced regarding the grade levels at which specific word analysis skills are introduced" (p. 95).

The content validity of the oral reading section has been established in reference to basal reading series rather than school curricula. This fact, in addition to the stringent criterion for mastery (97%), results in a demand for unrealistically high levels of decoding skill for some of the passages, particularly at the lower levels. For these reasons, caution should be exercised in interpreting results.

The manual cautions, with regard to language dominance screening, that "the results of the assessments in both languages should be compared and cross-validated with other data to identify the student's primary language" (p. xix).

With regard to the optional scoring section of each student screening form, the manual urges that if this scoring system is employed, the test results should be compared with other supporting data, such as local criteria in order to rank the test items.

The Examiner's manual provides potentially useful appendices, particularly a chart of "New York State and Federal Requirements" that can be met by using the Brigance in Spanish. Also provided are specific task correlations of this battery with the Brigance Comprehensive Inventory of Basic Skills.

Although the manual includes instructions for determining a student's primary language, caution should be observed in making such decisions. Test results should be used with other information about the students.

Reference

Manual

Publisher

Curriculum Associates

Title	<u>Dos Amigos Verbal Language Scales</u>	Grade	<u>K-6th</u>
Author	<u>D. Critchlow</u>	Time	<u>Untimed (approximately 20 minutes)</u>
Copyright	<u>1974</u>	Type	<u>NRT</u>

Intended Purpose of Test

The Dos Amigos Verbal Language Scales are intended to identify the student's dominant language, reveal the comparative development of the student's English and Spanish, and yield the level at which the student's spoken language development best qualifies him/her to function.

Description

This is a norm referenced, diagnostic instrument that consists of lists of Spanish and English words administered individually. As the administrator reads each word, the student responds by stating an acceptable antonym. After five consecutive errors in one language, the scale is administered in the second language. It yields percentile scores by chronological age.

Norm Data

It was standardized in Texas on 1,224 Mexican-American children in grades kindergarten to 6, whose first language was Spanish.

Reliability

The author reported test-retest reliability coefficients of .70 and .62 for kindergarten students and .97 and .90 for students in grade 1.

Validity

The author claimed concurrent validity based on positive, though undocumented, relationships between scores of 30 children on this instrument and on the Gilmore Oral Reading Test.

Special Alerts/Comments

The Dos Amigos provides an indication of oral language dominance. The results may be used to obtain supplementary, descriptive information with respect to a student's oral language dominance in Spanish or English. Given the limited standardization sample, the results are insufficient for determining a student's dominant language. This test was standardized on Mexican-American students, not on the Hispanic population in New York City. The administrator should be proficient in Spanish and English.

It has been suggested that the test underestimates Spanish dominance (Erickson & Omark, 1981). Caution should be taken when interpreting test scores, since performance may be influenced by dialectal variations.

The results of this test may not be used to determine if the student should be assessed by a bilingual or a monolingual team. In New York City, the Language Assessment Battery (LAB) is the only instrument used for this purpose.

References

Manual
Pietcher, et al, 1978
Erickson & Omark, 1981

Publisher

Academic Therapy Publications

Title	<u>Language Assessment Battery (LAB-82)</u>	Grade	<u>K - 12th</u>
Author	<u>M. M. Abbott</u>	Time	<u>Approximately 68 minutes</u>
Copyright Date	<u>1982</u>	Type	<u>NRT</u>

Intended Purpose of Test

The LAB-82 evaluates proficiency in listening, speaking, reading, and writing in English and Spanish. The LAB-82 is designed to determine entitlement to bilingual and ESL programs for students whose first language is not English. It is also used as the exit criterion for students in either program.

The Spanish LAB-82 evaluates Spanish language ability in listening, speaking, reading, and writing.

Description

The LAB-82 consists of two forms, four levels for each form, in both English and Spanish. The Spanish version, which is not a translation, was developed concurrently with the English edition and closely parallels it. Level I is used for grades kindergarten, 1, and 2; Level II, grades 3, 4, and 5; Level III, grades 6, 7, and 8; Level IV, grades 9, 10, 11, and 12.

Each level of the LAB-82 includes subtests of listening, speaking, reading, and writing. In Level I, listening and speaking are combined in a single subtest and reading and writing are in another subtest.

All subtests, except speaking, are group administered and the items are in multiple-choice format. In the individually administered speaking test, individual responses are scored separately for relevance and for grammar. The speaking subtest has no norms for levels, II, III, and IV.

Norm Data

The selection of items for LAB-82 was based upon data generated from an extensive pilot test program conducted during the 1981-82 school year. The sample for the pilot test program represented all grade levels of the New York City public school system.

Norms have been developed for subtest and total scores for both forms of the tests, and for Spring and Fall administrations; however, no norms exist for the speaking subtest in Levels II, III and IV. Students obtain a raw score (total of correct answers) which converts into a percentile score.

Reliability

Form A reliability studies were conducted discretely by grades for both the English and Spanish language versions of LAB-82. These grade reliabilities for limited English proficient (LEP) students on the English version for

individual subtests ranged from .80 to .95 with 80% above .87. The total test coefficients ranged from .92 to .97. For the Spanish version, grace reliability for individual subtests ranged from .85 to .97 with 80% above .88. The total test coefficients ranged from .90 to .98.

Validity

Content validity was assured by selection of relevant curriculum objectives and development of appropriate items to measure the identified objectives. Evidence that the battery is sensitive to different levels of and changes in English language proficiency was provided by data that reflect changes in item difficulty from Fall-to-Spring and Spring-to-Spring test administrations.

Special Alerts/Comments

The English LAB-82 measures English language ability in listening, speaking, reading, and writing. The Spanish LAB-82 measures Spanish language ability in listening, speaking, reading, and writing.

The English LAB-82 is used to determine student entitlement to Bilingual/ESL services. In New York City, a student, for whom English is a second language, who scores at or below the 20th percentile on the English version is entitled to bilingual or ESL service. In special education; a student who scores at or below the 20th percentile is entitled to be evaluated by a bilingual team. For specific uses in Special Education, see the Standard Operations Procedures Manual Appendix I: Provision of Services for other than English Language Students (SOPM ESL/Bilingual, 8/85).

A Spanish-speaking student who is entitled is given the Spanish version. The comparison between the English and Spanish scores is used to determine the type of services for Special Education students (See SOPM ESL/Bilingual, 8/85).

The only objective of the LAB-82 is to determine entitlement to Bilingual/ESL services. Percentile scores may be reported for LEP and/or bilingual students. It should not be used to diagnose levels of second language acquisition or reading, writing, speaking, and listening strengths and weaknesses. Comparison between subtests is not encouraged since the data do not support meaningful interpretations and the information obtained is of questionable value.

The LAB-82 should not be used for students for whom English is not a second language.

References

LAB-82- norms booklet
M.M. Abbott, 1985

Publisher

New York City Public Schools

Title	<u>Language Assessment Scales I and II (LAS)</u>	Grade	<u>LAS I: K-5</u> <u>LAS II: 6 and up</u>
Author	<u>E. De Avila & S. Duncan</u>	Time	<u>20 minutes</u>
Copyright Date	<u>I: 1977, II: 1978</u>	Type	<u>NRT</u>

Intended Purpose of Test

The LAS is designed to assess the student's "linguistic proficiency" in four areas: phonemic, lexical, syntactical, and pragmatic in both Spanish and English.

Description

The LAS, in both English and Spanish versions, consists of an examiner's kit, pictorial stimuli, an audio cassette, and score sheets. Language stimuli are provided to elicit responses which demonstrate "proficiency" along the authors' four parameters. Students respond orally or by pointing. The test is administered individually and production is scored at one of five proficiency levels which are unrelated to age.

Norm Data

The LAS I was standardized on 107 Spanish-speaking students, ages 5 to 12, of Mexican American, Latin American, and Puerto Rico backgrounds.

The LAS II was standardized on 311 English speaking students and 229 Spanish speaking students, ages 12 to 19, residing throughout the U.S. A second sample included 31 English speakers between the ages of 12 and 17 in a middle class California suburb. A third sample consisted of 34 Spanish speakers, ages 12 to 18, from an urban Latin American Site.

Results are reported as raw scores and total scores which may be converted into proficiency levels. The comparison between English and Spanish levels provides relative linguistic proficiency (RLP).

Reliability

High interrater reliability coefficients ranging from .84 to .98 were reported for both LAS I and LAS II (Manual).

Validity

Mann Whitney U test results with a small sample indicated significant differences ($p < .001$) between English speaking and limited English speaking groups when tested on the LAS I (Manual). On the LAS II, ANOVA results obtained with a small sample indicated significant differences ($p < .001$) between English and Spanish dominant groups on all subtests and total score (Manual).

Special Alerts/Comments

The LAS provides an indication of language proficiency in Spanish and English. The results may be used to obtain supplementary, descriptive information with respect to a student's language proficiency in Spanish and English. Given its technical limitations, the data are insufficient for determining a student's level of language proficiency in English or Spanish. The administrator must be fluent in both languages. The results of this test may not be used to determine if the student should be assessed by a bilingual or a monolingual team. The Language Achievement Battery (LAB) is the only instrument used for this purpose in New York City.

References

Manual
Pletcher, et al, 1978

Publisher

Linguametrics Group

Title	<u>La Prueba Riverside</u> <u>de Realización en Español-</u> <u>Form A (La Prueba)</u>	Grade	<u>K - 8th</u>
Author	<u>N. Cole , E. Trent, &</u> <u>D. Wadell</u>	Time	<u>30 - 60 minutes</u>
Copyright Date	<u>1984</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

La Prueba is designed to evaluate the skills of Spanish-speaking students in five content areas: Reading, Language, Mathematics, Social Studies, and Science.

Description

La Prueba consists of five subtests that are administered individually or to a group in a multiple choice format. There are nine levels, 6 through 14, which cover grades kindergarten through 8. In the lower levels, 6 to 9, students report answers directly in the test booklets which are either machine or hand-scored. Levels 10 to 14 have separate answer grids.

The Reading and Mathematics subtests have been translated and adapted into Spanish from the 3-R's Test. The Social Studies and Science subtests were created specifically for this battery. Students in grades kindergarten and 1 are tested only in Reading and Mathematics. Students in grade 2 are tested additionally in Science and Social Studies. Students in grades 3 through 8 may be tested in all five content areas.

Norm Data

The standardization sample consisted of 6,513 students in grades kindergarten through 8 in Texas. Results are reported as percentiles, normal curve equivalents, and stanines.

The authors pointed out that development of national norms was considered inappropriate, because of the difficulty in defining a national population that is representative of Spanish-literate students. La Prueba has been designed to allow users to develop their own reference norms.

Reliability

Internal consistency obtained by the Kuder-Richardson Formula 20 for individual subtests ranged from .70 to .92.

Validity

Content validity was indicated through consistency with Texas elementary school social studies and science curricula and confirmed by curriculum specialists at the University of Iowa. Subtest intercorrelations were also calculated; they ranged from .47 to .91.

Special Alerts/Comments

La Prueba is a battery of tests, given in Spanish, that are designed to evaluate skills in Reading, Language, Mathematics, Social Studies, and Science. Given its technical limitations, specifically the composition of the standardization sample, reliability estimates, and content based on Texas and not New York City curricula, the data are insufficient for making screening, performance level, instructional planning, or evaluation of student progress decisions.

Scores should not be reported for LEP and/or bilingual students.

The Prueba may also be categorized as a criterion referenced test since it provides information on a student's strengths and weaknesses. With other measures, La Prueba may be useful for developing shortterm objectives on the IEP.

Users of the test should be aware of regional variations in Spanish vocabulary and pronunciation. There is some evidence of social and family stereotyping in the Social Studies subtest. At times, illustrations are too small to be interpreted accurately. The degree of difficulty of the Reading subtest for Levels 8 to 14 is high for the norming population, as revealed by Prueba scores.

Reference

Manual

Publisher

Riverside Publishing Co.

Title	<u>Screening Test of Spanish Grammar (STSG)</u>	Age	<u>3 - 7 years</u>
Author	<u>A.S. Toronto</u>	Time	<u>Untimed</u> <u>(15 minutes)</u>
Copyright Date	<u>1973</u>	Type	<u>NRT</u>

Intended Purpose of Test

The STSG is designed to identify Spanish-speaking students who do not demonstrate native syntactic proficiency commensurate with their age.

Description

Individually administered, the STSG consists of subtests which measure receptive and expressive language. Twenty-three syntactic structures are assessed. The structures are presented in paired sentences which generally differ only in a single syntactic element. The items are arranged in order of increasing difficulty.

The test is too short to establish a basal and ceiling level; the complete test should be administered. If the student misses all items on the first half of the subtest, testing for that section should be discontinued.

Norm Data

The test was standardized on 192 students, 96 of Mexican and 96 of Puerto Rican descent between the ages of three and six years, eleven months. For each ethnic group there were 24 subjects, 12 males and 12 females in each year group, with Spanish as the primary language; all had been in the U.S. for a minimum of two years and were judged as having normal hearing and normal intelligence.

Reliability

Test-retest coefficients were computed using Spearman's rank order correlation procedure for matched groups ($N=24$); receptive language coefficients were .69 for the Mexican group and .62 for the Puerto Rico group; expressive language coefficients were .91 and .72, respectively.

Validity

Validity was based on correlation coefficients from the Developmental Assessment of Spanish Grammar and the STSG. Coefficients for receptive language were .55 and .63 for Mexican and Puerto Rican samples, respectively. For expressive language, coefficients were .64 for both samples.

Special Alerts/Comments

The STSG provides an indication of expressive and receptive language in Spanish. The results may be used to obtain supplementary, descriptive information with respect to language proficiency in Spanish. Given its technical inadequacies, specifically size and composition of the sample, the data are insufficient for determining a student's level of language proficiency in Spanish.

Scores should not be reported for LEP and/or bilingual students.

Reference

Publisher

Manual

Northwestern University Press

Title	<u>Secondary Level English Proficiency Test (SLEP)</u>	Grade	<u>7th - 11th</u>
Author	<u>Educational Testing Service</u>	Time	<u>Section 1: 45 minutes Section 2: 45 minutes</u>
Copyright Date	<u>1981</u>	Type	<u>NRT</u>

Intended Purpose of Test

The SLEP measures ability in understanding spoken and written English for the purpose of helping to make placement decisions related to the following: (1) assignment to English as a Second Language (ESL) class, (2) placement in a mainstream English language program, (3) exemption from a bilingual program, and (4) exit from an ESL program.

Description

The SLEP consists of 150 multiple-choice questions, divided into two subtests: during the listening comprehension subtest, the student listens to recorded passages in English and responds to written questions; the reading comprehension subtest is based on written and pictorial materials measuring vocabulary, grammar and overall reading comprehension. Stencils are provided to hand score the test. The content is reflective of US culture and its high schools. The material for the test was based on curricula typically found in American classrooms (grades 7 to 11). The context for some items, particularly the conversations used in the Listening Comprehension section, includes situations representative of those encountered by students in American secondary schools.

Norm Data

Prior to the final assembly of the pretest, 1,500 questions were written and reviewed by ETS staff and secondary school ESL specialists. Six thousand secondary school students in the U.S. and a number of foreign countries participated in the pretesting of the items. The results indicated that 84% of the pretested items could be used; two operational forms were developed with them.

One form of the SLEP was administered twice throughout the world; data from these administrations provided initial information about the statistical characteristics of the test. To gather data about the LEP population, 1,700 LEP students were given the test at more than 50 U.S. secondary schools during 1980-81.

The norms were established for separate categories of students by combining test data with information about individual student characteristics such as years of prior English study and length of enrollment in the current school.

Reliability

Indices of internal consistency, obtained through the Kuder-Richardson Formula 20, were high. Total test reliability was reported as .96; however, the highest subtest coefficient was .89 for the cloze section.

Validity

The intercorrelation of subtests ($r = .78$) suggests that each section measures aspects of language acquisition that are closely related but not identical (Stansfield, 1984). In addition, the low to moderate correlations between test scores and length of enrollment in the current school support the criterion-related validity of the SLEP (Stansfield, 1984).

Special Alerts/ Comments

The SLEP measures ability to understand spoken and written English for students learning English as a second language in grades 7 to 11. While its main purpose is to determine entrance and exit criteria to ESL and exemption from bilingual education classes, it cannot be used in New York City for those purposes due to current legal mandates.

The SLEP may be used to obtain additional information on student ability to understand spoken and written English, but may not be used to make decisions on language of assessment for clinical evaluation nor for placement. With other measures, the SLEP may be useful for developing shortterm objectives on the IEP in the area of ESL.

Scores should not be reported for LEP and/or bilingual students.

The results of this test may not be used to determine if the student should be assessed by a bilingual or a monolingual team. In New York City, the Language Achievement Battery (LAB) is the only instrument for this purpose.

References

Manual
Stansfield, 1984

Publisher

Educational Testing Service

Title	<u>Spanish Language Assessment Procedures (SLAP)</u>	Age	<u>5 - 8 years</u>
Author	<u>L.J. Mattes</u>	Time	<u>Untimed</u>
Copyright Date	<u>1985</u>	Type	<u>CRT</u>

Intended Purpose of Test

The SLAP is designed to assess structural and functional aspects of communication in Spanish.

Description

The SLAP consists of three components assessing a variety of listening and speaking skills. Part 1, Criterion Referenced Assessment of Spanish Communication, assesses 26 basic communication skills. Part 2, Assessment of Spanish Speech Sound Production, tests the student's ability to articulate 18 consonants in words and in sentences; the test is designed for students who demonstrate articulation errors in conversational speech. Part 3, Communication Sampling and the Identification of Communicative Disorders, consists of a language sample for the analysis of the student's utterances.

Test Administration

The SLAP is administered individually by a trained speech/language professional.

Technical Information

The individual tasks in this inventory were field-tested on Spanish-speaking student's in public school setting and have been reviewed extensively by professionals in both bilingual education and special education programs.

Effective Use/Comments

The SLAP provides information to assess structural and functional aspects of communication which may be used for descriptive reporting of a student's ability in the Spanish language. Given the inadequate technical information, results may not be used for making screening, performance level, instructional planning, and evaluation of student progress decisions. The information may be useful for developing shortterm objectives on the IEP in the area of native language arts.

Specific task items may need to be changed if the Spanish dialect used in the local community differs from the dialect used in the inventory. It is recommended that the inventory be reviewed by native Spanish speakers from the local community so that necessary modifications can be made in the content of the individual tasks.

The assessment tasks described in the manual should be administered by fluent speakers of the Spanish language who have the literacy skills necessary to present these tasks and to record the student's responses.

Reference

Manual

Publisher

College Hill Press

Title	<u>Test de Vocabulario en Imágenes</u> <u>Peabody - Adaptación</u> <u>Hispanoamericana (TVIP)</u>	Age	<u>2 1/2 to 18 years</u>
Author	<u>L. Dunn, E.R. Padilla, D.E. Lugo,</u> <u>& L.M. Dunn</u>	Time	<u>Untimed (10-15 mins.)</u>
Copyright Date	<u>1986</u>	Type	<u>NRT</u>

Intended Purpose of Test

The TVIP is designed to measure an individuals receptive vocabulary in Spanish. It is an adaption of the PPVT-R in English.

Description

The TVIP, which is individually administered, includes 125 vocabulary items arranged in increasing order of difficulty. Each item consists of four pictures presented on a page in an easel format. The examiner reads a statement or a word for each item and the student points to the picture which best represents the word.

Norm Data

Standardization was based on two samples. One set of norms consisted of 1,219 students, ages 2 1/2 to 16 from public schools in the Federal District of Mexico. Females and males were equally represented in the sample. All students came from urban and suburban areas. No data on SES were available.

The second sample consisted of 1,488 children ages 2 1/2 to 18 from public and private schools in San Juan, and in one small town. Females and males were equally represented. A majority of the sample (73.8%) lived in the metropolitan area and (26.9%) lived in the outlying areas. There was wide variability in proportionate representation of SES grouping across age levels.

A set combined norms, established for general use, was developed from a single pooled distribution of raw scores from the two above samples.

Results are reported as raw scores, age equivalents, standard scores, stanines, percentiles, and deciles.

Reliability

Measures of interval consistency were calculated at all age levels. The reliability coefficients are very high for all ages except for 2-year-olds ($r = .80$); for ages 3 and above, the coefficients ranged from .91 to .95.

Validity

Content validity was established as a result of test construction and item selection. Vocabulary items from the original item pool for the English PPVT were selected to cover 18 categories ensuring a breadth of vocabulary. The plates and words were field tested and refined. The final pool of 350 items was translated into Spanish and reordered in order of difficulty based on the Rasch-Wright item calibration procedures. Further subjective and empirical methods were used to ensure that the 125 stimulus words were universal Spanish.

Few studies have been conducted with the TVIP directly. Modest correlations were reported from limited research with the TVIP and the Kaufman Assessment Battery for Children - Spanish and the Habilidad General.

Special Alerts/Comments

The TVIP is designed to measure an individual's receptive vocabulary in Spanish which may be useful for making screening, performance level, instructional planning, and evaluation of student progress decisions in the area of receptive vocabulary. While the manual suggested a wide number of uses for this test, it emphasized that in most cases the recommended use of the scale is strictly as an achievement test. In addition, it warned test users to avoid over generalization from a screening device that measures only hearing vocabulary. Results are sufficient for developing longterm objectives on the IEP in the area of communication and language.

The results of this test may not be used to determine if the student should be assessed by a bilingual or a monolingual team. In New York City, the Language Assessment Battery (LAB) is the only instrument used for this purpose.

Reference

Manual

Publisher

American Guidance Service, Inc.

Title	Wechsler Intelligence Scale for Adults in Spanish (EIWA)	Age	16-64 years
Author	D. Wechsler	Time	1 hour
Copyright	1968	Type	NPT

Intended Purpose of Test

Like the WAIS in English, the EIWA is designed to evaluate intellectual functioning.

Description

The EIWA is a translation, adaptation, and standardization of the WAIS for the Puerto Rican adult population. It consists of materials comparable to the WAIS.

Norm Data

The standardization sample was comprised of 1,127 adults, 16 to 64 years old, residing in rural and urban areas of Puerto Rico.

Reliability

The reported split-half reliabilities ranged from .95 to .98 for the three scales (Verbal, Performance, and Total) for adults, ages 16 to 19. Test-retest reliability for one subtest, digit symbol, was reported as above .82 for this same age group.

Validity

The EIWA was based directly on the WAIS. Although primarily a translation of the WAIS, the content was changed to remove irrelevant items, add appropriate questions, and maintain levels of difficulty.

Special Alerts/Comments

The EIWA provides an indication of general intelligence of native Puerto Rican adults. The test findings may be useful for determining the level of intellectual/cognitive functioning of students at least 16 years of age who were raised in Puerto Rico and have recently arrived in New York City. Similar to the WAIS, the EIWA may only be administered, scored, and interpreted by a school psychologist.

Since the EIWA was translated and adapted in 1967 for an adult population from Puerto Rico, the test may be used to obtain descriptive information with respect to intellectual/cognitive functioning of older students who come from other Spanish speaking countries and older students who have been raised in New York City.

Scores should not be reported for LEP and/or bilingual students.

Reference

Manual

282

Publisher

Psychological Corp.

Title	<u>Wechsler Intelligence Scale for Children</u> <u>in Spanish (EIWN - R)</u>	Age	<u>5-15 years</u>
Author	<u>D. Wechsler</u>	Time	<u>1 hour</u>
Copyright	<u>1982</u>	Type	<u>NRT</u>

Intended Purpose of Test

Like the WISC-R, the Spanish version is designed to measure the overall capacity of an individual student to understand and cope with the world around him/her.

Description

It is a translation and adaptation of the WISC-R into Spanish for Hispanic students. It is used in conjunction with the WISC-R manual.

Norm Data

No norms have been developed as yet for use in New York City.

Reliability

No data are available for this translated edition.

Validity

The WISC correlated .67 with the Goodenough Draw-A-Person Test. There was a discrepancy between the norms of the Puerto Rican adaptation and those of the original scale. Although validity and order of difficulty of each test and items were determined, the distribution curve of IQ's on the Puerto Rican WISC was not a normal one. The Puerto Rican WISC mean IQ was 88.01 and the median 87.94; the standard deviation was 21.60. The Puerto Rican mean WISC IQ was approximately 12 points lower than the American one.

Special Alerts/Comments

The EIWN-R provides a measure of the overall capacity of an individual student to understand and cope with the world around him/her. It may be used to obtain descriptive information with respect to intellectual/cognitive functioning of Spanish speaking students in New York City. Given its technical inadequacies, the results of the EIWN-R are insufficient for determining a level of intellectual/cognitive functioning for Spanish speaking students in New York City.

Similar to the EIWN-R, the WISC in Spanish may only be administered, scored, and interpreted by a school psychologist.

Scores should not be reported for LEP and/or bilingual students. A norming project which was initiated in 1986 by the Office of Educational Assessment will be completed in 1987. Norms, reliability and validity information for this test will be available in September 1988; at that time the new norms may be reported as a level of intellectual/cognitive functioning.

Reference

Manual

Publisher

Psychological Corp.

Title	<u>Woodcock Language Proficiency Battery, Spanish Form (WLPB-Span)</u>	Grade	<u>K to 12</u>
Author	<u>R. W. Woodcock</u>	Time	<u>Untimed (45 minutes)</u>
Copyright	<u>1981</u>	Type	<u>NRT</u>

Intended Purpose of Test

The WLPB-Span is designed to measure language proficiency in Spanish in oral language, reading, and writing. It represents selected portions of the W-J Battery.

Description

The WLPI-Span contains eight subtests: three measures of oral language ability, three measures of reading abilities in Spanish, and two subtests which measure written language skills. Included in the battery are test records, stimulus material on easel format, and the key for correct choices.

Norm Data

General Spanish norms, obtained from January to July, 1980, were based on a sample of 802 subjects, grades Kindergarten to 12 from typical urban areas in Costa Rica, Mexico, Peru, Puerto Rico, and Spain.

Data were not provided for Spanish-speaking persons in the United States; the author stated that such norms would be invalid and questionable.

Results are reported as percentile ranks, standard scores, grade equivalents, normal curve equivalents, and the relative performance index.

Reliability

Internal consistency for individual subtests determined by split-half procedures, ranged from .77 to .90. The four cluster reliabilities were all above .88.

Validity

Content validity was developed as a result of item selection and test construction. Concurrent validity was indicated through correlations of the WLPB-Span and the Language Assessment Scales, Ber-Sil Spanish Test, and Dailey Language Facility Test. Construct validity was demonstrated by the low to moderate correlations between subtests which indicate that the subtests are measuring different abilities.

Special Alerts/Comments

The WLPS-Span provides a measure of language proficiency in Spanish with respect to oral language, reading, and writing ability. Based upon the standardization sample, the information provided is insufficient for making performance level decisions on Spanish language achievement. The information provided is sufficient for making evaluation of student progress and instructional planning decisions and developing short- and longterm objectives on the IEP in the area of Spanish language proficiency.

Because of the number of manipulations required in scoring, the administrator should be trained in the full use of the test. This test should be used in conjunction with the English edition, reviewed in the Communication/Language Skills section of the Guide.

Scores should not be reported for LEP and/or bilingual students.

Reference

Manual
Technical Summary

Publisher

DLM Teaching Resources

VOCATIONAL ASSESSMENT

Vocational assessment consists of collecting information on abilities, aptitudes, achievement, and interests. Most of the instruments included in this section are interest inventories. With tests reported in other sections of the Guide, these instruments provide information that will assist students in reaching decisions on courses of study and occupations.

Other instruments in this section include the measurement of knowledge and skills in specific vocations, attitude scales, and observation checklists.

Title	<u>APTICOM</u>	Grade	<u>H.S. and above</u>
Author	<u>Vocational Research Institute</u>	Time	<u>90 minutes</u>
Copyright Date	<u>1985</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

APTICOM is a microcomputer system for assessing an individual's aptitude, job interests, and language and math skills for the purposes of vocational guidance, training, and/or job placement.

Description

APTICOM is administered individually through the use of a computer terminal with the assistance of a paraprofessional. This test consists of three batteries in which input and output are effected by means of a microcomputer. The Aptitude Test Battery consists of 11 subtests; the Occupational Interest Inventory consists of 12 subtests; and the Educational Skills Development Battery consists of two subtests, Language and Math, which are subdivided into four levels. Combinations of results are classified according to 66 Occupational Aptitude Patterns and Work Groups, which are defined in the Guide for Occupational Exploration.

Two computer printouts may be obtained; a detailed Comprehensive Score Report or a Summary Score Report. Both reports list test results for the three batteries and provide vocational recommendations, but the Comprehensive Score Report also includes profiles of standard scores and detailed subtest analyses. The authors suggested the use of the Comprehensive Score Report for counselling purposes and the Summary Score Report as an "efficient record". The test kit includes the System Technical Manual, APTICOM System Operation, and Administration and Scoring Manual.

Norm Data

The initial standardization sample for the Occupational Interest Inventory consisted of 573 subjects, who were drawn from five sites across the country. The sample was stratified by sex, ethnicity, and number of years of education. Most of the high school students were in vocational programs. Renorming research samples (1985), based on 1,381 subjects, were divided into two groups, prevocational and vocational. Each group was drawn from approximately eight sites located throughout the country.

The Aptitude Test was field tested as a paper and pencil test version at several sites and the standardized administration procedures were validated with 31 vocational students from Philadelphia. A minimum performance level for the Educational Skills Development Battery was established through comparisons with other criterion referenced tests.

Test results on the Aptitude Battery and Interest Inventory are reported as raw scores, standard scores, and percentiles which are presented on a student profile. Score reports in the Educational Battery are analyzed according to "passing" minimum performance levels.

Reliability

Test-retest reliability estimates for the Aptitude Battery, based on a 1-week interval, ranged from .65 to .89. Internal consistency for the Interest Inventory, based on high school students, was high. Test-retest reliability estimates ranged from .74 to .86. There was no reliability information reported for the Educational Battery.

Validity

No validity information was reported for the Educational Battery. Concurrent validity, indicated a strong relationship between the Aptitude Battery and the GATB with only four of the 11 subtests. The factor structures of these two tests however, were reportedly dissimilar. Construct validity for the Interest Inventory was developed through an indication of the relationship of scores to the USES Interest Inventory.

Special Alerts/Comments

Although APTICOM provides information about an individual's aptitude, job interests, and language and math skills for the purposes of vocational guidance, training, and/or job placement, the information provided is not sufficient for making screening, performance level, instructional planning, or evaluation of progress decisions in any of the above areas.

As an interest inventory, the results are insufficient for determining a level of vocational development. With other measures, it may be useful for making individual instructional planning decisions. By itself, the Occupational Interest Inventory the Aptitude Test Battery, and the Educational Skills Development Battery provide information for developing shortterm objectives on the IEP in the area of vocational education.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously. Scores may not be reported for bilingual and/or LEP students.

Reference

Manual

Publisher

Vocational Research Institute

Title	<u>Assessment of Basic Vocation Related Skills (ABV-RS)</u>	Grade	<u>Secondary-Post Secondary</u>
Author	<u>C. Cline, N. Hartley, & K. Otazo</u>	Time	<u>45 minutes</u>
Copyright Date	<u>1979</u>	Type	<u>CRT</u>

Intended Purpose of Test

The ABV-RS is designed to provide information on skills and interests of student's in vocational education programs.

Description

The ABV-RS consists of 16 modules covering specific vocational areas: welding, sheet metal, agriculture, graphic arts, woodworking, automotive, autobody, drafting, health occupations, world of work, food preparation, child care, distributive education, electricity, horticulture, and business and office occupations. Each module contains materials for assessing academic and motor skills, and vocational skills, interest, and awareness as follows:

Reading Skills	Two paragraphs (grade levels 4-8 and 9-12) with vocabulary and comprehension questions.
Listening Skills	Written responses to cloze passages read to the student two times.
Locating Information in a Book	Using the table of contents and index.
Mathematics	Solving computation and word problems.
Motor Skills	Checklist for estimating: manipulative abilities including use of large and small tools, small motor-eye-hand and two-hand coordination, finger dexterity, steady work pace, and heavy lifting.
Daily Living Competencies	Checklist for estimating abilities in money concepts, transportation, job seeking skills, work completion, understanding directional terms, and safety rules, cooperation, punctuality and attendance.
Personal Characteristics	Checklist on grooming, personal hygiene, physical tolerance for full workday, peer relationships, self-confidence, and ability to work alone or in a group.

Work Aptitudes

Checklist on following
written/verbal instructions,
demonstration, multi-stepped
instruction, retaining instructions,
improving with repetition,
conceptualization of the problem,
and judgment

Vocational Interest/Awareness

Questionnaire

The World of Work includes two additional areas: Following Directions and Safety Warnings Quiz.

Pages in each packet are color coded for ease of usage and efficiency.

Test Administration

Specific directions are provided for each subtest in each of the packets. Motor Skills, Daily Living Competencies, Personal Characteristics and Work Aptitudes are evaluated on a 5-point scale through examiner observation. Subtests requiring direct student performances include the four academic skill areas; Following Directions, the Safety Warning Quiz, and the vocational awareness and interest questionnaire, each of which may be administered individually or to a small group. No special training or special materials are required.

Technical Information

The ABV-RS was developed to provide teachers with an informal assessment of the needs of students in vocational programs. It is based on the assumption that vocational programs can be modified to meet the individual needs of disadvantaged and handicapped students, but that such modification is most effective when directly related to assessed learning needs.

Field testing took place in Colorado in the Spring of 1979. It consisted of an initial review, reworking, classroom use, and workshop review. The materials were reviewed and/or used by vocational and special needs educators at each of the four testing stages.

The ABV-RS has not been normed; reliability and validity information are not available.

Effective Use/Comments

The ABV-RS was developed to assess the skills and interests of students in vocational programs. The information provided is not sufficient for making screening, performance level, instructional planning, or evaluation of student progress decisions given the method of assessment and limited

technical information and standardization. With other tests, the information is sufficient for making instructional planning decisions and for developing long- and shortterm objectives on the IEP in the area of career exploration.

Although the ABV-RS may be useful for obtaining descriptive information about the student in reading, listening, and mathematics, the test findings are insufficient for making individual instructional planning decisions in these areas.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

Reference

Manual

Publisher

Vocational Teacher Educational

Title	<u>Career Assessment Inventory (CAI)</u>	Age	<u>Adolescents and Adults</u>
Author	<u>C.B. Johansson</u>	Time	<u>30-40 Minutes</u>
Copyright Date	<u>1982</u>	Type	<u>NRT</u>

Intended Purpose of Test

The CAI is designed to identify an individual's vocational interest for career exploration, planning, and selection. Based on Holland's vocational model, its focus is on careers requiring up to four years of college.

Description

The CAI consists of 305 multiple choice items which are written at an 8th grade reading level. Test directions are clear and concise. Students can complete the inventory in 30 to 40 minutes independently. Those students who are unable to read the items may have the test read aloud. Completed answer documents are sent to NCS Interpretive Scoring Systems for scoring; a profile or interpretive report is available.

Information is reported on 16 General Interest Themes, 22 Basic Interest areas, and 91 Occupational Scales. Basic Interest Areas and Occupational Scales are referenced to the General Themes which are; realistic, investigative, artistic, social, enterprising, and conventional.

The CAI may be administered individually or to a group of students.

Norm Data

The standardization sample was reported as 1500 adults, half male and half female. No information was presented on age, geographic distribution, minority representation, or handicapping conditions.

Reliability

Correlation coefficients resulting from test-retest comparisons over 7-day to 30-day intervals were above .90 for all scales. Test-retest coefficients, reported for intervals exceeding six years, were above .80 for general themes and occupational scales and above .77 for basic interest area scales.

Validity

In addition to content validity, cross validation procedures resulted in empirical data indicating that respondents in specific occupational groups scored appropriately high or low on each scale.

Special Alerts/Comments

The CAI provides an indication of vocational interests which may be useful in obtaining supplementary, descriptive information. Given the standardization sample, the results are insufficient for determining a level of vocational development. With other measures, it may be useful for making instructional planning decisions.

The information provided by the CAI is sufficient for developing shortterm objectives on the IEP in the area of vocational education for students who read above fourth grade level.

Spanish and French-Canadian versions are available. However, the use of the CAI with culturally and linguistically diverse students and LEP students has not been demonstrated. Scores should not be reported for LEP and/or bilingual students.

Reference

Manual

Publisher's

NCS Interpretive Scoring Systems

Title	<u>Career Assessment Inventories For Learning Disabled (CAI-LD)</u>	Age	<u>Elementary School-Adult</u>
Author	<u>C. Weller & M. Buchanan</u>	Time	<u>20-30 minutes</u>
Copyright Date	<u>1983</u>	Type	<u>CRT</u>

Intended Purpose of Test

The CAI-LD is designed to facilitate career exploration and career choice among learning disabled students through an understanding of student strengths and deficits.

Description

Three inventories are included in CAI-LD:

Attributes Inventory	assesses the individual's dominant personality characteristics (realistic, investigative, artistic, social, enterprising, conventional).
Ability Inventory	assesses the individual's ability in the auditory, visual, and motor areas.
Interest Inventory	estimates student interests and assesses compatibility of interests with personality attributes and abilities.

The information derived from the inventories allows the teacher, counselor, or psychologist to identify appropriate career options from the Job Finder. The Job Finder, based upon Holland's theories of careers, presents job information in relation to personality characteristics and abilities.

Test Administration

The Attribute and Ability Inventories are completed by a teacher or counselor for one student at a time. The authors noted that "after one to three hours of observation or interview with the individual, the examiner should have little difficulty rating an individual on the Attributes and Ability Inventories" (Manual, p.17). The Interest Inventory is filled out by the student, individually or in groups. The material may be read aloud to the student.

Technical Information

Technical information is limited to a brief discussion of the theoretical basis of CAI-LD. Reliability is not addressed.

Effective Use/Comments

The information provided in the CAI-LD is sufficient for developing short term objectives on the IEP in the area of vocational education. With other measures, the findings may be useful for making instructional planning decisions. Based upon the technical information, results are insufficient for making performance level decisions about a student's vocational development.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

Reference

Publisher

Manual

Academic Therapy Publications

Title	<u>CDM: Career Decision-Making System (CDM)</u>	Grade	<u>7th - Adult</u>
Author	<u>T.F. Harrington & A.J. O'Shea</u>	Time	<u>25-50 minutes</u>
Copyright Date	<u>1982</u>	Type	<u>NRT</u>

Intended Purpose of Test

CDM is designed to identify an individual's vocational interest for career exploration, planning, and selecting.

Description

The CDM System consists of 120 statements to which a student responds on a 3-point scale (LIKE, CAN'T MAKE UP YOUR MIND, DISLIKE) indicating degree of interest in career clusters. Scores are tabulated and compared to the other areas. A chart of 18 Career Clusters may be used to explore specific job opportunities.

The CDM can be administered to a group of students and can be self-or machine-scored. Both English and Spanish versions are available. However, a student who reads below a 7th grade level should have the test read aloud. Students reading below a 4th grade level may have difficulty with the concepts.

Norm Data

The CDM was administered to 9,650 students in grades 7 through 12, in 32 school districts, stratified by enrollment and socioeconomic status, and 27 colleges in the United States during 1981. Occupational clusters, identified for each student, are based on raw scores. Separate scales are available for males and females as well as combined sex scales.

The Spanish edition was field tested and revised among 648 Spanish-speaking subjects, grades 7 through adult, from Puerto Rico, Texas, California, and Massachusetts. The sample was stratified by sex.

Reliability

Internal consistency coefficients ranged from .91 to .94 for the standardization sample of junior high and high school students. Studies of test-retest reliability with high school students reported results typical for interest inventories; reliability estimates were generally between .78 and .90. Comparisons of student scored and author scored protocols were exceedingly high ($r = .96$ and $.99$, respectively).

Validity

The authors reported that the CDM has construct and concurrent validity (Buros, 1985). The CDM was based upon Holland's model of vocational development.

Concurrent validity was based on differences in scale scores of each occupational group surveyed. The authors obtained evidence of the CDM's predictive validity.

Special Alerts/Comments

The CDM provides an indication of vocational interest which may be useful in obtaining supplementary, descriptive information. As an interest inventory, the results are insufficient for determining a level of vocational development. With other resources, it may be useful for making instructional planning decisions and for developing short- and longterm objectives on the IEP in the area of vocational education for students who read above fourth grade level.

Although an edition is available in Spanish, which was standardized on a bilingual population in the United States and for which validation studies indicated as positive results as in the English edition, it should be used cautiously with culturally and linguistically diverse students and LEP students. Scores from the Spanish version may not be reported for Hispanic LEP and/or bilingual students.

References

Manual
Buros, 1985

Publisher

American Guidance Service

Title	<u>Career Occupational Preference System (COPS)</u>	Grade	<u>JHS to Adult</u>
Author	<u>R. R. Knapp & L. Knapp</u>	Time	<u>Untimed (30 minutes)</u>
Copyright Date	<u>1984</u>	Type	<u>NRT</u>

Intended Purpose of Test

COPS is an interest inventory designed to measure interests related to occupations and careers in each of 14 clusters: Science-Professional/Skilled; Technology-Professional/Skilled; Consumer Economics; Outdoor; Business-Professional/Skilled; Clerical; Communication, Arts-Professional/Skilled; and Service-Professional/Skilled.

Description

COPS consists of 168 job activity statements to which a student responds on a 4-point scale. Responses may be marked directly in the individual test booklet or on an answer sheet. It is self or group administered and may be hand- or machine-scored.

A Spanish translation is available. A student who reads below a 6th grade level should have the test read aloud. Students reading below a 4th grade level may have difficulty with the concepts.

Norm Data

The 1982 standardization sample was based on testing 6,834 students in grades 7 through 12, from five regions in the country. The sample included representative proportions of urban, rural and minority groups according to the 1981 census.

Results are reported as stanines and percentiles.

Reliability

Alpha reliability coefficients for the 14 career clusters ranged from .86 to .91. Test-retest reliability based on retesting 97 tenth graders over a 1-week interval ranged from .80 to .91; based on retesting 497 seventh graders over a 1-year interval, coefficients ranged from .62 to .80.

The stability of interest measurement was examined by retesting 1,243 students, grades 7 through 12 over a 1-year interval; results revealed 45% agreement between major cluster choices.

Validity

Construct validity was indicated through correlations of COPS with similar inventories; coefficients ranged from .50 to .93.

Concurrent validity, indicated by comparing COPS to work values, choice of college major, and occupations in the military revealed moderate to high percentages of agreement. Predictive validity was demonstrated through follow-up studies.

Special Alerts/Comments

COPS provides an indication of vocational interest which may be useful in obtaining supplementary, descriptive information. As an interest inventory, the results are insufficient for determining a level of vocational development. Combined with other measures, it may be useful for making instructional planning decisions and developing short- and longterm objectives on the IEP in the area of vocational education for students who read above fourth grade level.

Results with linguistically and culturally diverse students and LEP student should be interpreted cautiously. Scores should not be reported for LEP and/or bilingual students.

References

Manual
Technical manual

Publisher

Educational and Industrial
Testing Service

Title	<u>Career Orientation Placement and Evaluation Survey (COPES)</u>	Grade	<u>7th - 12th</u>
Author	<u>R. R. Knapp & L. Knapp</u>	Time	<u>20 - 50 minutes</u>
Copyright Date	<u>1978</u>	Type	<u>NRT, CRT</u>

Intended Purpose of Test

COPES is designed to help students become aware of work values related to career planning. Eight dimensions are measured: Investigative, Practical, Independence, Leadership, Orderliness, Recognition, Aesthetic, and Social.

Description

COPES consists of 2-choice paired comparison statements that reflect work values. There are 160 statements on the Machine-Scoring Form and 128 statements on the Self-Scoring Form. The inventory is self-administered and responses may be marked directly in the individual test booklets. The test may be administered in small groups or individually. Each item may be read aloud.

Work values of the COPES may be compared with interests on the COPS and interpreted in relation to the occupational clusters of the COPSsystem.

Norm Data

The standardization sample was based on testing 6,253 students in grades 7 through 12, as well as 700 community college students. Means from all sample groups were combined to form a single norms sample. No details regarding race, geographic area, or socioeconomic level were provided in the manual. Results are reported as percentiles.

Reliability

Internal consistency reliability coefficients ranged from .67 to .81 and were considered to represent "an acceptable level of reliability commensurate with other value inventories". There was no mention of test-retest reliability data in the manual.

Validity

Construct validity of COPES scales is based on factor analysis undertaken to determine the structure of work values; intercorrelations of the dimensions were low and supported their independence. Content validity was indicated through consistency with popular taxonomies of value dimension. Predictive validity was demonstrated through follow-up studies.

Special Alerts/Comments

COPES is designed to help students become aware of their work values related to career planning. The results are insufficient for

determining a level of vocational development. With other resources, it may be useful for making instructional planning decisions for students who read above fourth grade level.

Positive features are the ease of administration and scoring and interpretive usefulness. COPES may also be categorized as a criterion referenced test since it provides information on a student's strengths and weaknesses.

Due to the COPES high verbal loadings, it should be used cautiously with linguistically and culturally diverse students and LEP students. Scores should not be reported for LEP and/or bilingual students.

Reference

Manual

Publisher

EDITS

Title	<u>Florida International Diagnostic Prescriptive Vocational Competency Profile (Florida)</u>	Age	<u>Adolescent - Adult</u>
Author	<u>H. Rosenberg & D.G. Tesolowski</u>	Time	<u>Untimed</u>
Copyright Date	<u>1979</u>	Type	<u>CRT</u>

Intended Purpose of Test

The Florida is designed to measure vocational competency in six areas for handicapped adolescents and adults.

Description

The Florida consists of a 70-item checklist divided into six vocational competency domains to measure an individual student's vocational competency. The six areas are: Vocational Self-Help Skills, Social-Emotional Adjustment, Work-Attitudes-Responsibility, Cognitive-Learning Ability, Perceptual-Motor Skills, and General Work Habits. Each item is rated on a 5-point scale ranging from ready for placement in competitive employment to unable to perform task. Raw scores for each domain and a composite score are entered in the Record Form, converted into percentages, and recorded on the Profile of Strengths and Weaknesses. Individualized diagnostic and prescriptive forms are also included.

Test Administration

The Florida is completed by an evaluator who is familiar with the individual being tested. Ratings are based on observed behavior relating to vocational competency. The authors recommended the use of the instrument following observation of the individual in all six vocational competency domains.

Technical Information

Content validity was established through test development procedures; items were selected from an extensive literature search and interviews with school personnel.

The sample, divided into secondary and postsecondary levels, was stratified by age, sex, race, and time in training program.

Interrater reliability and concurrent validity were established through testing 100 handicapped trainees between the ages of 13 and 65. Interrater reliability coefficients ranged from .74 to .90; concurrent validity estimates ranged from .76 to .96.

Effective Use/Comments

The Florida is designed to measure vocational competency of handicapped adolescents and adults. However, as with other third-party rating scales, the Florida is not a direct measure of student performance and, consequently, judgment of vocational competency requires additional direct behavioral

information. With other measures, the Florida may be useful for individual evaluation, selection and placement, program/training evaluation, and establishing shortterm objectives for IEPs in the area of vocational education. The instrument provides insufficient information for determining performance level decisions in vocational education. While objectives are specific, strategies for obtaining these objectives are not provided.

Since it focuses on observation of behavior, the Florida may be used effectively with linguistically and culturally diverse students and with LEP students.

Reference

Manual

Publisher

Stoelting Co.

Title	<u>Interest Determination, Exploration, and Assessment System (IDEAS)</u>	Grade	<u>6th-12th</u>
Author	<u>C.B. Johansson</u>	Time	<u>30-40 minutes</u>
Copyright Date	<u>1983</u>	Type	<u>NRT</u>

Intended Purpose of Test

IDEAS is designed to identify student vocational interests for career exploration and planning. IDEAS was developed from the Career Assessment Inventory (CAI).

Description

IDEAS consists of 112 multiple choice items, written at a sixth grade reading level to which a student responds on a 5-point scale. Test directions are clear and concise and include information on self-scoring procedures. IDEAS may be administered individually or to a group of students. Students can complete IDEAS independently including reading the directions, responding to items, and scoring and plotting scores. An easy-to-read profile displays the individual's scores graphically. IDEAS may be read aloud to students whose reading level is below sixth grade.

Fourteen interest areas are measured: mechanical/fixing, electronics, nature/outdoors, science, numbers, writing, arts/crafts, social service, child care, medical service, business/sales office practices, and food service. Each scale is referenced to careers in the Occupational Outlook Handbook and the Dictionary of Occupational Titles. In addition, suggested courses relating to each scale are included.

Norm Data

The standardization sample was based on 598 male and female students in grades six through eight and 3436 male and female students in grades nine through twelve. No information was presented on the geographic distribution of students, minority group representation, or whether students had handicapping conditions. Separate norms are presented for males and females.

Reliability

Test-retest reliability coefficients were reported in the high 80's and low 90's.

Validity

High content validity was indicated through test development procedures. Concurrent validity was demonstrated by the correlation of individuals in specific occupations and career programs with a related test scale score.. Construct validity was established through comparisons with similarly used scales.

Special Alerts/Comments

IDEAS provides an indication of vocational interests which may be useful in obtaining supplementary, descriptive information. Given the standardization sample, the results are insufficient for determining a level of vocational development. With other measures, it may be useful for making instructional planning decisions.

The information provided by IDEAS is sufficient for developing short- and longterm objectives on the IEP in the area of vocational education. Its use with culturally and linguistically diverse students and LEP students has not been demonstrated. Scores should not be reported for LEP and/or bilingual students.

References

Manual
Buros, 1985

Publisher

NCS Interpretive Scoring Systems

Title <u>Job Awareness Inventory (JAI)</u>	Grade <u>10th-12th</u>
Author <u>T. Makowski</u>	Time <u>Untimed</u>
Copyright Date <u>1980</u>	Type <u>CRT</u>

Intended Purpose of Test

The JAI is designed to assess average and special needs students in prevocational and career education skills.

Description

The JAI is available in two forms. Part A contains 85 questions and may be used as a pretest or survey to determine how much the student knows about the world of work; Part B, consisting of 75 questions, additionally serves as a posttest, measuring accomplishment.

Each form is divided into four parts, as follows:

I Occupations	multiple choice items identifying occupations
II Do you Know How To	multiple choice items about job search activities
III General Information	multiple choice items on personnel information and other job-related matters
IV Interview Actions	forced choice items on behavior during an interview

Test Administration

The JAI may be administered individually or in a group. No special training or materials are required.

Technical Information

The JAI was field tested in six high schools for two years. Students who were classified as educable mentally retarded, learning disabled, or deaf were also included in the field test.

For a sample of 13 students in grades 10 to 12, retested after a 10-day interval, reliability was estimated as .86. Validity and score interpretation were not addressed.

Effective Use/Comments

The JAI assesses students in the areas of prevocational and career education skills. The information provided is sufficient for developing short- and longterm objectives on the IEP. Instructional strategies are not included, but the author refers to instructional programs produced by the publisher.

The test is applicable to Special Education students. However, based upon the technical information, the results are insufficient for making performance level decisions about a student's vocational development.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

Reference

Publisher

Manual

Mafex Assoc., Inc.

Title	<u>JOB-0</u>	Grade	<u>7th and above</u>
Author	<u>A. Cutler, F. Ferry, R. Kouk, & R. Robinett</u>	Time	<u>Untimed (45 minutes)</u>
Copyright Date	<u>1978</u>	Type	<u>CRT</u>

Intended Purpose of Test

JOB-0 is designed to assess vocational interests of students at all educational levels. It provides the student with an opportunity for self-awareness, self-assessment, career exploration, and decision-making.

Description

JOB-0 consists of a reusable assessment booklet and consumable answer folders on which the student indicates interests and compares these interests with those of persons who are in specific job titles. Nine variables are assessed and are related to 120 job titles. A JOB-0 Dictionary describing each job title is part of the test materials.

Careers are explored in relation to the following contents or variables:

Educational Expectations, Activities, Interpersonal Relationships, Leadership, Physical Activity.

Test Administration

JOB-0 may be administered individually or to a group. Students respond in folders to nine questions and compare their responses to those of persons in the job titles. The examiner requires no special training.

Technical Information

The test is an outgrowth of the Dictionary of Occupational Titles, the Occupational Outlook Handbook, and FIRO-B. It was piloted on 20,000 students in grades nine through twelve.

Reliability was reported as standard errors of measurement for each variable for seven occupational groups. Low estimates of standard errors of measurement were obtained from responses of a small group representing each group.

Effective Use/Comments

JOB-0 provides an indication of vocational interest of students at the intermediate and high school level. As an interest inventory, results are insufficient for determining a level of vocational development or for making individual instructional planning decisions. It may be useful in an instructional program of career awareness and exploration of students reading above fourth grade level.

JOB-O software is not recommended at this time, until the programming is revised.

JOB-O is available in Spanish. Its specific use with Special Education students and with culturally and linguistically diverse students and LEP students has not been demonstrated.

Reference

Manual

Publisher

CFKR Career Materials

Title	<u>Occupational Aptitude Survey and Interest Schedule-Interest Schedule (OASIS-IS)</u>	Grade	<u>Jr. and Sr. High School</u>
Author	<u>R. M. Parker</u>	Time	<u>35 - 45 minutes</u>
Copyright Date	<u>1983</u>	Type	<u>NRT</u>

Intended Purpose of Test

OASIS-IS is designed to assess the interests of junior high school and high school students for vocational exploration and career development.

Description

OASIS-IS consists of 240 items scored on a 3-point scale of like, neutral, and dislike. Each item is assigned to one of twelve subscales: Artistic, Scientific, Nature, Protective, Mechanical, Industrial, Business Detail, Selling, Accommodating, Humanitarian, Leading-Influencing, and Physical Performing.

Norm Data

The standardization of OASIS-IS was based on a sample of 1,398 nonhandicapped students in grades 8 through 12 from eleven states. There were equal numbers of males and females represented. White students from urban areas predominated.

Results are reported as raw scores, percentiles, and stanines.

Reliability

Test-retest reliability based on a 2-week interval with 54 junior high and high school students was relatively high; subscales ranged from .66 to .91. Inter-item consistency or alpha reliability was high; the median coefficients for the 12 scales ranged from .86 to .94.

Validity

Content validity was demonstrated by the relationship of OASIS-IS items to the job activities presented in the 1979 Guide for Occupational Titles from the Department of Labor.

Special Alerts/Comments

The OASIS-IS is designed to assess the interests of junior high school and high school students for vocational exploration and career development. Due to the limited standardization sample, item content, and technical

information, the information provided may be useful for obtaining supplemental information about the student in the area of vocational education. With other measures, the findings may be useful for developing longterm goals.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously. Scores should not be reported for LEP and/or bilingual students.

Reference

Manual

Publisher

Pro-Ed

Title Occ-U-Sort
Author L. K. Jones
Copyright Date 1981

Grade 8th - Adult
Time 60 minutes
Type CRT

Intended Purpose of Test

Occ-U-Sort is designed to gather information on a student's vocational interests. It may be used as a career counseling tool, to encourage career exploration, and to assist in the decision-making process about future occupations.

Description

Occ-U-Sort consists of a deck of 2"x 3" cards, a self-guided booklet, and a guide to occupations. On the front of each card is the (1) occupational title, (2) the Dictionary of Occupational Titles (DOT) number, (3) the estimated educational level determined by the Department of Labor, and (4) a two letter occupational code which references the occupation to Holland's system of six categories. On the back of the card is a short occupational description. The self-guided booklet permits a student to use the material independently by providing a step-by-step description of the process for selecting potential occupations. The occupational guide is a tool to increase the student's knowledge of possible occupational choices.

There are three decks of cards, each organized according to general education development (GED) estimates. Plus 3 cards contain occupations ranging from below a high school diploma through a graduate degree; two-thirds of the occupations in Plus 3 are within a high school level of education or require some additional training. Plus 4 cards include occupations requiring extended on-the-job training or some college. Plus 5 cards require a college education.

Test Administration

Occ-U-Sort may be self-administered or administered by a counselor, psychologist, or other professional on an individual basis. Group administration would require multiple decks of cards since one deck is necessary for each participant. By using a systematic procedure of five steps, the student is able to select occupations which are of interest. Two additional steps provide for the expansion of knowledge regarding career choices.

Technical Information

Occ-U-Sort is based upon the theories and findings of early researchers in career education, the DOT and the Department of Labor's supplement to the DOT, and Holland's 1973 material which organized occupations into the following six scales: realistic, investigative, artistic, social, enterprising, and conventional.

Many studies have been reported on the reliability and validity of Occ-U-Sort in facilitating career exploration and decision-making. The author indicated that Occ-U-Sort codes have been found to be moderately stable. Comparisons with other inventories such as the Vocational Preference Inventory indicated similarities of career selections. Studies have also shown that use of Occ-U-Sort broadened career options especially for females. Moreover, research comparing the effectiveness of expressed interests vs. inventoried interests demonstrated the equivalence of these two methods of assessing vocational interests.

No direct evidence of predictive validity exists at the present time.

Effective Use/Comments

Occ-U-Sort provides an indication of vocational interest of students in grade 8 or above, who read above fourth grade level. As an interest inventory, results are insufficient for determining a level of vocational development. With other measures, it may be useful for making individual instructional planning decisions.

The information provided by Occ-U-Sort is sufficient for developing short- and longterm objectives on the IEP in the area of vocational education. The instrument may be useful in an instructional program of career exploration and development.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

Reference

Manual

Publisher

Publisher's Test Service

Title	<u>Reading-Free Vocational Interest Inventory</u>	Age	<u>13 years - Adult</u>
Author	<u>R.L. Becker</u>	Time	<u>Untimed (20 minutes)</u>
Copyright Date	<u>1981</u>	Type	<u>NRT</u>

Intended Purpose of Test

The Reading-Free Vocational Interest Inventory is designed to provide systematic information on the vocational interest patterns of mentally retarded and learning disabled students and adults in semi-skilled and unskilled occupations.

Description

This pictorial vocational interest inventory contains no verbal symbols or written statements. Items are simple line drawings of vocationally related tasks. The examinee is asked to circle one picture from a set of three that he/she "likes best". Subjects are shown 55 triads which identify eleven interest areas. The test can be administered individually or in groups.

Norm Data

Norms were based on a nationwide sample from 30 states. Over 6,500 were educable mentally retarded in public secondary schools and sheltered workshops; over 4,000 were learning disabled. IQ's of the mentally retarded were in the 60's, whereas IQ's of the learning disabled were in the 90's. Separate norms are provided for the learning disabled and educable mentally retarded at different age levels for each sex.

Reliability

Test-retest reliability coefficients obtained from tests given over a 2-week period were in the .70's and .80's. Internal consistency was estimated by a Kuder-Richardson formula and ranged from .68 to .92, with a median of .82 for male groups in both residential and public school settings. The reliability for public school females ranged from .70 to .96, with a median of .81.

Validity

Content, concurrent, and occupational validity were cited. Items were examined by various study teams and subjected to statistical analysis to determine their discriminating power. Correlations between this test and similar interest areas on the Geist Picture Interest Inventory were generally low to moderate; however, the correlation coefficients were not given.

Special Alerts/Comments

The Reading-Free Vocational Interest Inventory provides an indication of vocational interest of mentally retarded and learning disabled students which may be useful in obtaining supplementary, descriptive information.

As an interest inventory, the results are insufficient for determining a level of vocational development. With other measures of vocational ability, it may be useful for making instructional planning decisions with mentally retarded and learning disabled students, especially those students in Basic II.

Given the standardization sample, the inventory should only be used with educable mentally retarded and learning disabled populations. Buross (1978) had noted that the 1975 edition depicted no minority group members; the new revised test consists of the same line-drawings as the earlier version. The major change in this new 1981 edition is the increase in a non-sexist approach to occupations and the inclusion of learning disabled students in the standardization sample.

Its use with culturally and linguistically diverse mentally retarded and learning disabled students and LEP mentally retarded and learning disabled students has not been demonstrated.

References

Manual
Buross, 1978

Publisher

E.P. Elbern Publications

Title San Francisco Vocational Competency
Scale (San Francisco)

Age Mentally Retarded
Adults - 18 or older

Author S. Levine & F.S. Elzey

Time Untimed

Copyright Date 1968

Type NRT, CRT

Intended Purpose Of Test

The San Francisco is designed to measure the vocational competency of individuals whose primary disability is mental retardation. It was developed to select mentally retarded individuals for training in semi-independent and sheltered workshops, to assess an individual's status, to judge growth over time, to study the effect of different training procedures, and to screen for placement in independent work situations.

Description

The San Francisco is a rating scale which is completed by a supervisor to measure actual performance and not presumptive ability. Each individual is rated at the level at which he or she characteristically performs vocational tasks.

There are 30 items on the SFVCS which measure motor skills, cognition responsibility, and social-emotional behavior. Each item in the scale contains four or five options which range from a low level of competence (1) to a high level of competence (4 or 5). Each option is expressed behaviorally so that the rater reports on actual behaviors which are characteristically observed.

Norm Date

Percentile scores have been developed for males and females. They were based on ratings of 344 males and 218 females in 45 workshops located throughout the U.S. Participant age range varied from 18 to more than 30; more than one third were 18 to 20 years of age. About two-thirds had up to eight years of schooling, and half had up to two years of workshop experience. Male and female norms were developed because the average scores for each group were statistically significant ($p < .01$).

Reliability

Estimates of internal consistency based upon scores from the standardization sample were exceedingly high ($r=.95$) for both males and females. Correlation of test-retest ratings after a 1-month interval for 54 individuals resulted in a stability coefficient of .85.

Validity

Test development procedures assured content and construct validity of the San Francisco in relation to measurement of vocational competency.

Special Alerts/Comments

The San Francisco provides an indication of vocational competency of mentally retarded adults which may be useful in obtaining information for screening, determination of performance, instructional planning, and evaluation of student progress. It may be useful for obtaining only supplementary or descriptive information about the vocational competency of younger students.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously. Scores should not be reported for LEP and/or bilingual students.

Reference

Manual

Publisher

The Psychological Corporation

Title	<u>Social and Prevocational Information Battery Revised (SPIB-R)</u>	Grade	<u>EMR Junior and Senior High School</u>
Author	<u>A. Halpern & L. Irvin</u>	Time	<u>25 minutes per subtest</u>
Copyright Date	<u>1986</u>	Type	<u>CRT</u>

Intended Purpose of Test

The SPIB-R is designed to measure skills and competencies important for community adjustment of higher functioning mentally retarded individuals.

Description

The SPIB-R, like the original 1975 SPIB, consists of nine subtests, predominantly in a true/false format.

The goals, subtests, and content descriptions follow:

<u>Goal</u>	<u>Subtest</u>	<u>Content</u>
Employability	Job Search Skills	types of jobs; job requirements; application forms; job sources; job assistance agencies; use of telephone; interview behavior; resumes; classified ads
	Job Related Behaviors	role of supervisor; job related communications; job completion; work relations with peers
Economic Self-Sufficiency	Banking	checking and savings accounts; deposits and withdrawals; check-writing and cashing; check records; interest; check safety
	Budgeting	time buying; borrowing; arithmetic skills for budgeting; overspending; credit; relationship between budgeting and savings; expenses; paying bills
	Purchasing Habits	comparative shopping; types of stores; newspaper ads; weight and cost; purchasing terminology; sales tax; quantity buying

Family Living	Home Management	maintenance and repair; safety; sanitary living conditions, food preparation and storage; laundry procedures; home appliances; protective and utility agencies
	Health Care	emergency care; common practices; use of medication; child care; prenatal care; importance of care; body temperature
Personal Habits	Hygiene and Grooming	regular health care; body cleanliness; consequences of poor health care; cleansing agents and grooming aids
Communications	Functional Signs	traffic and road; safety; emergency; general information

The content of the SPIB-R is similar to the SPIB-T; however, this form is more appropriate for higher functioning students. Hand-and machine-scorable booklets are available.

Test Administration

The SPIB-R may be administered individually or to small groups in three sessions of three subtests each; a proctor should be available for every ten students. Unlike the original SPIB which consisted of oval-shaped bubbles to denote selected answers, students are required to place an "x" on the word TRUE or FALSE.

Results are reported as percentage correct and percentiles.

Technical Information

The SPIB-R was based upon the original field testing of the 1975 SPIB, which consisted of 1,800 Oregon junior and senior high school EMR students. Norms, validity, and reliability data were obtained from 453 of the students who were 14 to 20 years of age and of mostly Caucasian background.

KR20 reliability estimates ranged from .65 to .82 for subtests and .94 for the total test battery. Test-retest correlations had similar estimates. Mastery levels had not been established.

The manual states that new standardization data for the SPIB-R will be gathered over the next several years.

Effective Use/Comments

The information provided in the SPIB-R is useful for screening, instructional planning, and evaluation of student progress decisions in the areas of employability, economic self-sufficiency, family living, personal habits, and functional sign reading. Although the information is sufficient for developing short- and longterm objectives on the IEP in the area of vocational education, the results are insufficient for determining a level of vocational development. The manual includes guidelines for linking SPIB-R results to instructional planning, IEP development, and a set of curriculum materials (SIL) that are available from the distributor of SPIB-R.

A Spanish translation of selected items from each subtest, produced for Project TRABAJO with permission of this publisher, is available from the Special Education Testing Unit.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

Reference

Manual

Publisher

CTB/McGraw Hill

Title Social and Prevocational Information
Battery-Form T (SPIB-T)

Age TMR and EMR Adolescents

Author I. Irvin, A. Halpern, & W. Reynolds

Time Untimed (10 - 20 minutes
per subtest)

Copyright Date 1979

Type CRT

Intended Purpose of Test

The SPIB-T is designed to measure skills and competencies important for community adjustment of lower functioning mentally retarded individuals.

Description

The content of the SPIB-T is similar to the SPIB-R; however, this form is more appropriate for a lower functioning population.

Materials are provided for determining test-taking ability. Students who do not obtain a specified minimum score on the pretest are not eligible to take the SPIB-T.

The SPIB-T consists of nine tests which are administered orally to individuals or small groups of students. They measure growth in long-range goals which are common to most programs for this population.

<u>Goal</u>	<u>Subtest</u>	<u>Content</u>
Employability	Job Search Skills	types of jobs; job requirements; application forms; job sources; use of telephone; interview behavior, job assistance agencies; resumes; classified ads; factors in selecting a job
	Job Related Behaviors	role of supervisor; job completion; work relations with peers; personal habits
Economic Self-Sufficiency	Banking	opening accounts; making deposits and withdrawals; writing and cashing checks; interest; usefulness of checking accounts; check recording and safety
	Budgeting	time buying; borrowing; arithmetic skills for budgeting; overspending; credit; types of expenses; paying bills

	Purchasing Habits	comparative shopping; types of stores; weight and cost; newspaper ads; general purchasing terminology; sales tax; quantity buying
Family Living	Home Management	maintenance and repair; safety; sanitary living conditions; food preparation and storage; laundry procedures; home appliances; protective and utility agencies
	Health Care	emergency care; common practices; use of medication; child care; prenatal care; knowledge of body temperature;
Personal Habits	Hygiene and Grooming	regular health care; body cleanliness; consequences of inadequate care; cleansing and grooming aids
Communications	Functional Signs	traffic and road; safety; emergency; general information

Test Administration

The SPIB-T may be administered individually or to groups no larger than two or three students. Test should be scheduled over a period of several days. Pretests are administered first and students who score above the minimum are then given SPIB-T. Results are reported as raw scores and percentage correct for each subtest and total battery.

Technical Information

The SPIB-T was field tested on 287 residents of group homes and 188 TMR students whom the authors described as "moderately retarded". Slightly more than half of each group was able to take the SPIB-T. Test administration generally consisted of two to three 1-hour sessions.

Reliability estimates of internal consistency computed for each homogeneous sample in the field test varied from .68 to .87 for each subtest. Total battery estimates were .96 for the residential sample and .95 for the TMR sample.

Concurrent validity was determined for the residential sample by comparing SPIB-T scores with scores obtained on a highly reliable and valid Behavior Rating Form. Validity coefficients obtained from a sample of 129 residents ranged from .60 to .75 among the nine subtests and total score, thereby demonstrating a strong relationship between the SPIB-T and the performance ratings.

Effective Use/Comments

The information provided by the SPIB-T is useful for screening, instructional planning, and evaluation of student progress decisions in the areas of employability, economic self-sufficiency, family living, personal habits, and functional sign reading. Although the information is sufficient for developing short- and longterm objectives on the IEP in the area of vocational education, the results are insufficient for determining a level of vocational development.

The SPIB-T may not be appropriate for low-level trainable mentally retarded students with visual, auditory, and/or perceptual difficulties (Doucette and Freedman, 1980).

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

References

Manual
Doucette & Freedman, 1980

Publisher

CTB/McGraw Hill

Title	<u>Tests for Everyday Living (TEL)</u>	Grade	<u>9th - 12th</u>
Author	<u>A. Halpern L. Irvin, & J. Landman</u>	Time	<u>20-30 minutes per subtest</u>
Copyright Date	<u>1979</u>	Type	<u>CRT</u>

Intended Purpose of Test

The TEL is designed to measure knowledge of life skills of senior high school students who may be remedial and learning disabled students, but not mentally retarded. Tasks include shopping, using bank accounts, managing one's personal finances, obtaining and keeping a job, and maintaining one's health.

Description

The TEL consists of seven subtests which are administered orally to individuals or groups of students. Items are in a multiple choice format and may be hand-scored. The subtests measure growth in the following Life Skill areas:

<u>Life Skill Area</u>	<u>Subtest</u>	<u>Content</u>
Employability	Job Search Skills	types of jobs; job requirements, job assistance agencies; job sources; classified ads; application forms; use of telephone; resources; interview behavior; job related terminology
	Job Related Behavior	work relations with supervisors and peers; changing jobs; job training; work permits; payroll procedures; job related forms; unions
	Purchasing Habits	comparative shopping; types of stores; newspaper ads; sales tax; cost saving methods; restaurant purchasing; returning merchandise; weight markings, consumer terminology; arithmetic skills for purchasing; sales receipts
	Banking	checking and savings accounts; deposits and withdrawals; check writing and cashing; check records; interest; other bank services; other financial institutions
Economic Self-sufficiency	Budgeting	making and following a budget; expenses; overspending, borrowing; relationship between budgeting and saving; credit; grace periods; arithmetic skills for budgeting; reading and understanding budgets

Family Living	Health Care	preventative health care practices; treatment for problems; emergency care; use of medication; child care; body temperature
	Home Management	maintenance and repair; safety; sanitary conditions; laundry procedures; home appliances; protective and utility agencies

Test Administration

Tests should be scheduled for administration in three sessions with two or three subtests per session. The TEL may be administered individually or in groups no larger than 20 students. A proctor should be available for every ten students.

Technical Information

The TEL was developed at the Research and Training Center in Mental Retardation at the University of Oregon. It was field tested on 1,150 junior and senior high school students in Oregon, California, and Alaska; the high school students were identified as regular or remedial. No special education students were included.

Content validity was established through content and item selection procedures which included an exhaustive review of literature on life skills education for adolescents. Reliability was not addressed. Mastery levels have not been established.

Effective Use/Comments

The information provided by the TEL may be useful for screening, instructional planning, and evaluation of student progress decisions in the areas of employability, economic self-sufficiency and family living for students who are not mentally retarded. The SPIB-R or SPIB-T are appropriate for populations who have mental retardation. The information is sufficient for developing short- and longterm objectives in the IEP in the area of vocational education. Given the limited standardization sample, the results are insufficient for determining a level of vocational development.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

Reference

Manual

Publisher

CTB/McGraw Hill

Title Vocational Competency Measures Grade Secondary and Post-secondary
(VCM)
Author American Institutes for Research Time Untimed
Copyright Date 1983 Type LRT

Intended Purpose of Test

VCM is designed to assess job readiness among students who have completed vocational training.

Description

VCM consists of 17 comprehensive and independent tests which evaluate student competency in a specific job area. They are:

- | | |
|-------------------------------|--------------------------------------|
| 1. Agricultural Chemicals | 10. Physical Therapist Assistant |
| - Applications Technician | 11. Custom Sewing |
| 2. Farm Equipment Mechanic | 12. Restaurant Service |
| 3. Computer Operator | (Waiter, Waitress, Cashier) |
| 4. Word Processing Specialist | 13. Electronics Technician |
| 5. Apparel Sales | 14. Water Treatment Technician |
| 6. Fabric Sales | 15. Waste water Treatment Technician |
| 7. Grocery Clerk | 16. Carpenter |
| 8. Hotel (Motel) Front Office | 17. Diesel Mechanic |
| 9. Dental Assistant | |

Each of the 17 tests includes three subtests: Job Information Test, Performance Tests, and the Work Habits Inventory. The Job Information Test, a multiple-choice test of job relevant knowledge, is available in two parallel forms. There are several Performance Tests, each requiring the student to perform activities expected at an entry-level position; The Work Habits Inventory includes three questionnaires and covers 17 work related areas, such as dependability, initiative, honesty, perseverance, and appearance.

Each subtest comes in booklet form. The Examiner's Manual provides administrative, scoring, and technical information.

Test Administration

Test selection is dependent upon the student's specific field of training. The Job Information Test was designed for group administration but may be administered individually. The Performance Test is individually administered and the first of two parts of the Work Habits Inventory is self-administered; the third part must be completed by the student's instructor.

There is no special training required on the part of the examiner, except for the Performance Tests which also require equipment and materials normally available in any vocational program.

Suggestions for administration to students with hearing/vision impairments, orthopedic handicaps, and limited English proficiency are provided in the manual. Braille or translated versions are recommended but not provided.

The test administrator must design and provide answer sheets for the Job Information Test since none are included. Results are reported as raw scores.

Technical Information

VCM was field tested across the country among students enrolled in high schools, vocational education centers, technical institutes and/or junior colleges. The number of students varied for each of the 17 tests (e.g., Restaurant Service was field tested on 238 students; Grocery Clerk on 167).

Parallel form reliability and internal consistency were moderately high and varied for each test. Content validity was indicated by VCM's relevance to industry/trade requirements based on employer ratings. Concurrent validity was demonstrated through the correlation between test scores and number of years of job experience and supervisor's subjective evaluations of job performance. Other criterion validity was demonstrated by correlations of test scores to training and correlations with course grades.

Effective Use/Comments

VCM is designed to assess job readiness of students who have completed vocational training and may be useful for making instructional planning, and evaluation of progress decisions and for developing short- and longterm objectives on the IEP in the area of vocational education. It is not sufficient for determining a level of performance.

Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously.

Reference

Manual

Publisher

American Association
for Vocational Instructional
Materials

Title	<u>Wide Range Interest-Opinion Test (WRIOT)</u>	Age <u>5 years - Adult</u>
Author	<u>J.F. Jastak & S. Jastak</u>	Time <u>40-60 minutes</u>
Copyright Date	<u>1979</u>	Type <u>NRT</u>

Intended Purpose of Test

The WRIOT is designed to measure interests and attitudes related to career/vocational planning for handicapped and nonhandicapped individuals.

Description

The WRIOT may be administered in groups or individually. The test consists of a spiral book or 35mm filmstrip containing 150 plates, each displaying a choice of three pictured activities. The individual indicates those activities most and least preferred on an answer sheet. Specific test items or "Key Items" correlate positively with females or males and form specific interest and attitude clusters. The eighteen interest clusters are: Art, Literature, Music, Drama, Sales, Management, Office Work, Personal Service, Protective Service, Social Science, Social Service, Biological Science, Physical Science, Number, Mechanics, Machine Operation, Outdoor, and Athletics. The seven attitude clusters are Sedentariness, Risk, Ambition, Skill Level, Sex Stereotype, Agreement, and Negative/Positive.

In addition to the picture book and answer sheets, the WRIOT kit also includes a manual, 48 overlay stencils, and a report form. The manual provides cluster descriptions, correlated job titles, interpretations, and sample case studies. The answer sheet may be hand-scored with the use of overlay stencils or machine-scored. A report form summarizes results in terms of a profile.

Norm Data

The standardization sample was based on approximately 9,000 subjects between the ages of five and adulthood and "were in no way restricted to any economic, intellectual, or racial populations." The authors stated that no attempt was made to obtain a representative national sampling. The sample was stratified by sex and included all ethnic groups and proportionate numbers of below average, average, and above average levels of intelligence.

Tests results were reported as raw scores and converted to T Scores.

Reliability

Test-retest reliability was high, ranging from .82 to .95. No other reliability measures were calculated.

Validity

Correlations to determine construct validity between the WRIOT and the Geist Picture Interest Inventory indicated a very slight relationship. No other validity was discussed.

Special Alerts/Comments

The WRIOT measures interests and attitudes related to career/vocational planning for handicapped and nonhandicapped individuals. As an interest inventory, the results are insufficient for determining a level of vocational development. With other measures, it may be useful for making instructional planning decisions and for consultation with counselors.

The information provided by the WRIOT is sufficient for developing short- and longterm objectives on the IEP in the area of vocational education.

Positive aspects are the care taken to eliminate sex stereotyping and the representation of minority groups. The pictorial format is useful with students with limited schooling and little or no reading ability. Alternate procedures are provided for the blind or persons with a low concentration level. However, it is a lengthy test.

Although many scoring masks must be used to develop a complete profile, scoring procedures are relatively simple.

Caution should be used in interpreting results due to the limited amount of research and technical data.

Caution should be observed in using the WRIOT with culturally and linguistically diverse students and LEP students.

References

Manual

Publisher

Jastak Associates Inc.

GLOSSARY

This glossary was designed to assist the nontest specialist in using the Test Resource Guide and tests. The terms are generally found in manuals and other literature on educational and psychological testing. Definitions are based upon the usage commonly found in textbooks and other literature in the field of measurement and assessment.

Achievement test. An achievement test measures the extent of learning generally after instruction has been given.

Aptitude test. Aptitude tests are designed to measure an individual's ability to learn or develop proficiency in a specific areas. This category includes tests of general academic ability or intelligence tests and tests which measure special abilities such as mechanical, music, art, etc.

Battery. A battery consists of a group of tests standardized on the same population; as a result, scores obtained from these tests are comparable.

This term is informally applied by clinicians to a group of tests that will be administered to one individual.

Correlation. Correlation is a statistical procedure to determine the relationship between two scores for the same individuals. The correlation coefficient also indicates the proportion of variance which the measures share in common and how accurately scores on one test are predicted from scores on the other.

Criterion. A criterion is the measure against which a test is evaluated, primarily for providing evidence of validity. A criterion measure is generally a test, but may be report card grades, teacher judgment, etc.

Criterion referenced test (CRT). A CRT provides information on the extent to which a student possesses specific skills or objectives in a specific content area.

Factor analysis. Factor analysis is a statistical procedure to provide information on the interrelationships or "factors" being measured on a test.

Norm data. Norms are meaningful scores that result from statistical analysis of the actual (raw) scores obtained from the standardization sample. Information about this first group who took the test should include number, geographic distribution, community size, ethnicity, sex, socioeconomic status, age, and/or grade. If appropriate, ability level and specific handicapping condition should be indicated.

Norms are frequently reported as grade equivalents, percentiles, standard scores, and stanines. Grade equivalents are averages of raw scores obtained by the standardization sample in each grade. A grade equivalent of 3.6 means that the student obtained the same score as a member of the

standardization sample in the sixth month of the third grade. A percentile is that point on the score scale at or below which fall a specific percent or cases: for example, the 50th percentile is that point on the score scale at or below which are 50% of the cases. A student at the 93rd percentile is scoring as high or higher than 93% of the students in the standardization sample.

A standard score is a score which interprets the raw scores so that they have meaning: these scores are based on the variability (standard deviation) of the scores of the standardization sample. A stanine is an example of a standard score which is based upon a 9-point scale.

Norm referenced test (NRT). An NRT provides information about an individual in relation to other individuals of similar age/grade. The standardization sample from whom the test norms are developed must be representative of the students for whom test scores are obtained.

Personality test. A personality test measures an individual's affective characteristics through the techniques of self-report, rating scales, projectives in which ambiguous stimuli are presented to the subject, and attitude inventories.

Projective technique. A projective technique is a method of personality study in which an individual's responses are obtained by means of nonstructured or ambiguous stimuli such as ink blots, pictures, unfinished sentences, etc. It is based on the assumption that the individual projects evidence of personality characteristics and organization which is scored and interpreted to yield a description of each individual's basic personality structure.

Reliability. Reliability is an estimate of score consistency, generally reported as a coefficient of correlation or the standard error of measurement. Through reliability, the test user determines how trustworthy or dependable the score obtained from a specific test is.

Reliability is estimated in terms of the test's internal consistency, stability, or equivalence. For internal consistency, a split-half reliability coefficient, corrected by the Spearman-Brown correction formula, is obtained by correlating the scores on one half the test with scores on the other half; split-half reliability is especially appropriate for "power" achievement and aptitude tests. Kuder-Richardson formulas are also used to provide estimates of internal consistency. To determine the consistency or stability of a test score over time, a test-retest reliability coefficient is obtained by correlating scores of the same individuals resulting from two different test administrations; stability coefficients are appropriate in aptitude testing. Alternate-form reliability or estimates of equivalence are obtained by correlating scores of the same individuals resulting from two different forms of the test; generally two different test administrations of short interval are characteristic of the procedures to obtain coefficients of equivalence.

As a coefficient approaches 1, the greater the reliability. Generally reliability coefficients of .90 or above indicate that a test may be used for decisions about individuals.

The standard error of measurement (SEM) is used to determine the dependability of an individual's score on a test. The SEM indicates the amount by which a score differs from the hypothetical true score. The larger the SEM, the less reliable the score. By adding and subtracting the SEM to a score, the test user obtains an estimate of the score range in which the individual is likely to fall.

Standardization sample. The original population taking a test and from whose raw scores standard and derived scores are determined is known as the standardization sample.

Standardized test. A standardized test provides for administration and scoring according to a set of specific rules. It is generally a timed test that is published commercially.

Validity. Validity indicates whether a test measures what it is intended to measure.

For achievement tests, content validity is most appropriate. Content validity indicates the representativeness of the test. Procedures to determine content validity include a comparison of curriculum or courses of study and test content and item analyses.

Criterion-related validity, estimated by the coefficient of correlation, indicates the extent to which test scores are related to a second measure, the criterion. Predictive validity refers to comparisons made with a criterion measured in the future, whereas concurrent validity refers to comparisons made with a criterion administered at the same time. Correlations ranging from .40 and above are considered evidence of strong validity.

Construct validity indicates the extent to which a test measures a trait or concept; factor analytic procedures including canonical correlation are statistical methods commonly used to provide evidence of construct validity on personality and attitude tests.

APPENDICES

APPENDIX A

THE LAW AND THE ASSESSMENT PROCESS-LEGAL GUIDELINES*

OVERVIEW

Public Law 94-142 and Part 200 of New York State Regulations ensure that:

1. Each handicapped child must receive a complete, individual evaluation;
2. Tests and evaluation procedures must not be racially or culturally discriminatory;
3. Judgements must be based on overall educational need and not solely on a single intelligence score;
4. Handicapped children must be re-evaluated every three years;
5. A variety of assessment data will be used in the IEP development;
6. Parent's and children's rights are protected and due process is guaranteed.

P.L. 94-142 addresses eight different areas of concern regarding assessment. Each aspect of an assessment program must comply with the law.

The table below lists the assessment functions with the specific requirements of the P.L. 94-142 regulations that pertain to them. Assessment teams may use the table as a checklist for compliance with P.L. 94-142 as they develop assessment reports.

ASSESSMENT FUNCTIONS AND P.L. 94-142

121a. 531 Preplacement evaluation.

- . Individual evaluation required prior to special education placement

Before any action is taken with respect to the initial placement of a handicapped child in a special education program, a full and individual evaluation of the child's educational needs must be conducted in accordance with the requirements of 121a.532.

This was initially printed from Writing Individual Assessment Reports in Special Education: A Resource Manual, National Association of State Directors of Special Education, (NASDSE), Washington, DC. 1978.

* Permission to reproduce these Regulations was obtained in a letter from the National Association of State Directors of Special Education, (NASDSE), Washington, DC.

PROTECTION IN EVALUATION PROCEDURES

- . Non-biased assessment procedures required
 - 12la.530 General.
 - (a) Each State educational agency shall insure that each public agency establishes and implements procedures which meet the requirements of 12la.530-12la.534.
 - (b) Testing and evaluation materials and procedures used for the purposes of evaluation and placement of handicapped children must be selected and administered so as not to be racially or culturally discriminatory.
 - (20 U.S.C. 1412(5)(C).)
- . Evaluation criteria
 - 12la.532 Evaluation procedures. State and local educational agencies shall insure, at a minimum, that:
 - (a) Tests and other evaluation materials:
 - (1) Are provided and administered in the child's native language or other mode of communication, unless it is clearly not feasible to do so;
 - (2) Have been validated for the specific purpose for which they are used; and
 - (3) Are administered by trained personnel in conformance with the instructions provided by their producer;
 - (b) Tests and other evaluation materials include those tailored to assess specific areas of educational need and not merely those which are designed to provide a single general intelligence quotient;
- . Native language requirement
- . Valid use of tests
- . Administered by trained personnel
- . Not IQ only

. Must measure aptitude or achievement, not just reflect impairment

(c) Tests are selected and administered so as best to ensure that when a test is administered to a child with impaired sensory, manual or speaking skills, the test results accurately reflect the child's aptitude or achievement level or whatever other factors the test purports to measure, rather than reflecting the child's impaired sensory, manual, or speaking skills (except where those skills are the factors which the test purports to measure);

. More than one procedure

(d) No single procedure is used as the sole criterion for determining an appropriate educational program for a child; and

. Evaluation team

(e) The evaluation is made by a multidisciplinary team or group of persons, including at least one teacher or other specialist with knowledge in the area of suspected disability.

. Comprehensive assessment

(f) The child is assessed in all areas related to the suspected disability, including, where appropriate, health, vision, hearing, social and emotional status, general intelligence, academic performance, communicative status, and motor abilities.

. Re-evaluation every three years

121a.534 Re-evaluation.

Each State and local educational agency shall insure:

(a) That each handicapped child's individualized education program is reviewed in accordance with 121a.340-121a.349 of Subpart C, and

- (b) That an evaluation of the child based on procedures which meet the requirements under 12la.532, is conducted every three years or more frequently if conditions warrant or if the child's parent or teacher requests an evaluation.
- . Parental consent required prior to evaluation
- 12la.500 Definitions of "consent," "evaluation," and "personally identifiable."
- As used in this part: "Consent" means that:
- (a) The parent has been fully informed of all information relevant to the activity for which consent is sought, in his or her native language, or other mode of communication;
- (b) The parent understands and agrees in writing to the carrying out of the activity for which his or her consent is sought, and the consent describes that activity and lists the records (if any) which will be released and to whom; and
- . Consent necessary only for individual evaluation
- (c) The parent understands that the granting of consent is voluntary on the part of the parent and may be revoked at any time.
- "Evaluation" means procedures used in accordance with 12la.530-12la.534 to determine whether a child is handicapped and the nature and extent of the special education and related services that the child needs. The term means procedures used selectively with an individual child and does not include basic tests administered to or procedures used with all children²⁵ⁿ school, grade, or class.



CITY SCHOOL DISTRICT OF THE CITY OF NEW YORK

Regulation of the Chancellor

Category: STUDENTS

No.: A-831

Subject: NON-DISCRIMINATORY STANDARDS AND
PROCEDURES FOR THE REFERRAL,
ASSESSMENT AND PLACEMENT OF STUDENTS
WHO MAY HAVE HANDICAPPING CONDITIONS.

Page: 1 of 12

Issued: 10/1/84

ABSTRACT

These regulations apply to all schools and programs throughout the city district. The Non-discriminatory Standards and Procedures stated here seek to ensure that the referral assessment and placement process for students who may have handicapping conditions occurs in a fair and non-biased manner. These procedures are designed to guide all school staff in planning for the appropriate services for students from the many diverse cultural, linguistic and ethnic groups that constitute the New York City school system.

1. DEFINITIONS

- 1.1 As used herein, "Major language groups" shall be defined to include those languages identified by the Chancellor based on a survey of home languages and the need for bilingual assessment and instruction. The Chancellor shall review periodically and revise by addition or deletion this definition as necessary.
- 1.2 As used herein, the phrase "Ethnic group" shall be defined to include Black (not Hispanic), White (not Hispanic), Hispanic, American Indian or Alaskan Native, Asian, or Pacific Islander.



Regulation of the Chancellor

Category: STUDENTS

No.: A-831

Subject: NON-DISCRIMINATORY STANDARDS AND
PROCEDURES FOR THE REFERRAL,
ASSESSMENT AND PLACEMENT OF STUDENTS
WHO MAY HAVE HANDICAPPING CONDITIONS.

Page: 2 of 12

Issued: 10/1/84

2. REFERRAL

- 2.1 Prior to initiating a referral, the school staff member shall investigate regular education alternatives as an attempt to resolve the problem. If indicated, regular education alternatives are the preferred intervention. Appropriate intervention strategies should be tried, monitored for change and documented. Applicable laws and regulations regarding referrals shall be followed.
- 2.2 The school staff member shall present information supporting the referral to the appropriate personnel.
- 2.3 The school staff member initiating the referral shall record a brief descriptive statement of the problem, including descriptions of the student's behavior as it relates to the statement of the problem identified in the referral, the frequency of the behavior, and, unless clearly unnecessary, the comparable behavior of other students in the classroom.

3. ASSESSMENT

- 3.1 The District Subcommittee on the Handicapped or its designee, shall inform parents of the process used to assess their child's educational needs in school. They shall be informed about:
 - a. who made the referral;
 - b. why the referral was made;
 - c. any objective documentation of the reasons for referral in language they can understand;
 - d. their rights to participate in the assessment and decision-making process of the team;



CITY SCHOOL DISTRICT OF THE CITY OF NEW YORK

Regulation of the Chancellor

Category: STUDENTS

No.: A-831

Subject: NON-DISCRIMINATORY STANDARDS AND
PROCEDURES FOR THE REFERRAL,
ASSESSMENT AND PLACEMENT OF STUDENTS
WHO MAY HAVE HANDICAPPING CONDITIONS.

Page: 3 of 12

Issued: 10/1/84

- e. their right to examine relevant records with respect to such assessment, including the instruments themselves and the procedures to be used;
 - f. their right to stop the assessment process at any time by written notice; and
 - g. The Board of Education's right to request an impartial hearing to proceed with the assessment.
- 3.2 The assessment process shall be discussed with parents in the language spoken by parents in the home. Translators shall be used when needed and their use noted in the case record. Assessment personnel should be sensitive to misunderstandings that might result from linguistic differences.
- 3.3 The Division of Special Education shall establish and distribute procedures to determine when a student needs to be interviewed in an other than English language.
- 3.4 When a student needs to be interviewed in an other than English language, the interview shall be conducted by assessment personnel fluent in that language. When such personnel are unavailable and efforts to obtain them have been documented, a qualified translator shall be used. When a translator is used, the use shall be documented in the case record.
- 3.5 Assessment procedures to be used shall be centrally reviewed and recommended by the Division of Special Education, and distributed to assessment and other relevant personnel.



Regulation of the Chancellor

Category: STUDENTS

No.: A-831

Subject: NON-DISCRIMINATORY STANDARDS AND
PROCEDURES FOR THE REFERRAL,
ASSESSMENT AND PLACEMENT OF STUDENTS
WHO MAY HAVE HANDICAPPING CONDITIONS.

Page: 4 of 12

Issued: 10/1/84

- 3.6 Assessment personnel shall assure that there is a descriptive statement in the assessment reports regarding the stated referral problem.
- 3.7 When the stated referral problem or the problem identified by assessment personnel is behavioral, the assessment reports shall include descriptions of the student's behavior as it relates to the problem, the frequency of the behavior, and, unless clearly unnecessary, the comparable behavior of other students in the classroom.
- 3.8 Assessments shall provide information clearly relevant to the stated referral problem or problem identified by assessment personnel. Assessment information shall be in a form useful for decisions that need to be made for the provision of instructional and related services, including all reviews and re-evaluations.
- 3.9 Assessment records shall include at least the following information: Date, times, setting, name and title of person(s) collecting the information.
- 3.10 Assessment personnel shall collect several types of information (such as school records, direct observations, test scores, and self reports).
- 3.11 Assessment personnel shall collect information which describes, in terms of observable behavior, areas of the student's strength and competence as well as areas of problems and weaknesses.
- 3.12 Assessment personnel shall collect information indicating the referred student's strengths and weaknesses from persons who interact frequently with the student.



Regulation of the Chancellor

Category: STUDENTS

No.: A-831

Subject: NON-DISCRIMINATORY STANDARDS AND
PROCEDURES FOR THE REFERRAL,
ASSESSMENT AND PLACEMENT OF STUDENTS
WHO MAY HAVE HANDICAPPING CONDITIONS.

Page: 5 of 12

Issued: 10/1/84

- 3.13 Assessment personnel shall collect information about the interaction between the target student and peer(s), teacher(s), and others whenever relevant to the problem which generated the referral. The collection of this information shall be a part of systematic observational procedures that are directly related to specific questions generated by the referral problem.
- 3.14 At least one member of the assessment team shall conduct at least one structured observation of the student in the primary setting. The team member shall describe the results of this observation in the record.
- 3.15 When the presenting problem is behavioral, the assessment team is encouraged, consistent with their professional judgment, to conduct any or all of the following observations:
- a) More than one structured observation of the student in the primary setting;
 - b) At least one structured observation of the student in at least one setting other than the primary setting (e.g. other subject class, gym, lunch room);
 - c) Structured observations in the primary setting by at least two assessment team members.

If the presenting problem is behavioral and the assessment team conducts only one structured observation of the student in the primary setting, the assessment team shall state in the record the reason for not conducting additional observations.



CITY SCHOOL DISTRICT OF THE CITY OF NEW YORK

Regulation of the Chancellor

Category: STUDENTS

No.: A-831

Subject: NON-DISCRIMINATORY STANDARDS AND
PROCEDURES FOR THE REFERRAL,
ASSESSMENT AND PLACEMENT OF STUDENTS
WHO MAY HAVE HANDICAPPING CONDITIONS.

Page: 6 of 12

Issued: 10/1/84

- 3.16 Assessment personnel shall not employ assessment procedures beyond their stated purpose. If an assessment procedure is used beyond its stated purpose, the assessment personnel shall state why there are extraordinarily compelling circumstances which require its use and shall provide a written description of how the procedure was used and shall include that in the student's folder.
- 3.17 When selecting assessment instruments or procedures, whether in cognitive, emotional, or behavior areas, assessment personnel shall vigorously scrutinize available norms regarding their appropriateness. For standardized tests, if appropriate norms are not available, the information collected shall be of a descriptive nature.
- 3.18 When standardized tests or structured observations are employed, standard procedures must be followed.
- 3.19 Projective tests may provide descriptive information relevant in answering questions about the student's fantasy, motivation and reality testing. However, they should be used cautiously and with sensitivity to the individual student's linguistic, cultural and ethnic background.
- 3.20 Assessment personnel shall collect the referred student's self-reports of their thoughts and feelings in response to natural situations or standard stimuli, where appropriate, but extreme caution shall be exercised in their interpretation.



CITY SCHOOL DISTRICT OF THE CITY OF NEW YORK

Regulation of the Chancellor

Category: STUDENTS

No.: A-831

Subject: NON-DISCRIMINATORY STANDARDS AND
PROCEDURES FOR THE REFERRAL,
ASSESSMENT AND PLACEMENT OF STUDENTS
WHO MAY HAVE HANDICAPPING CONDITIONS.

Page: 7 of 12

Issued: 10/1/84

- 3.21 Assessment personnel shall collect information to rule out to the extent possible sensory or cognitive difficulties as the primary explanation of the stated referral problem or problem identified by assessment personnel.
- 3.22 Assessment personnel shall collect information to rule out to the extent possible linguistic, cultural or ethnic differences as the primary explanation of the stated referral problem or problem identified by assessment personnel.
- 3.23 Where appropriate, assessment personnel shall make recommendations for intervention which shall be tried, evaluated and documented in the present placement before completion of the assessment, if they are feasible within applicable time limits.



Regulation of the Chancellor

Category: STUDENTS

No.: A-831

Subject: NON-DISCRIMINATORY STANDARDS AND
PROCEDURES FOR THE REFERRAL,
ASSESSMENT AND PLACEMENT OF STUDENTS
WHO MAY HAVE HANDICAPPING CONDITIONS.

Page: 8 of 12

Issued: 10/1/84

4. CLASSIFICATION AND PLACEMENT

- 4.1 On request by parents, the School Based Support Team (SBST) or the District Subcommittee on the Handicapped shall provide copies of the written reports (psychological, educational evaluation, social history, etc.) before the Educational Planning Conference and/or District Subcommittee on the Handicapped review.
- 4.2 The SBST or the District Subcommittee on the Handicapped shall provide parents with copies of the Individualized Education Program and all required notices.
- 4.3 The SBST or the District Subcommittee on the Handicapped shall provide parents of limited English proficiency who speak major languages with written translations of the Individualized Education Program and all required notices. On request of the parent, oral translations of the assessment reports shall be provided.
- 4.4 The SBST or the District Subcommittee on the Handicapped shall inform parents of their right to have an advocate present at the Education Planning Conference or District Subcommittee on the Handicapped Review.
- 4.5 The SBST or the District Subcommittee on the Handicapped shall not consider a case without having documented that the student's parents were informed of their rights to participate, to have an advocate accompany them at the meeting and an interpreter present if necessary.



CITY SCHOOL DISTRICT OF THE CITY OF NEW YORK

Regulation of the Chancellor

Category: STUDENTS

No.: A-831

Subject: NON-DISCRIMINATORY STANDARDS AND
PROCEDURES FOR THE REFERRAL,
ASSESSMENT AND PLACEMENT OF STUDENTS
WHO MAY HAVE HANDICAPPING CONDITIONS.

Page: 9 of 12

Issued: 10/1/84

- 4.6 A member of the assessment team shall be present when the District Subcommittee on the Handicapped meets to review assessment information when there is a disagreement between the SBST team and the parent or upon request of the parent, a member of the assessment team or a member of the District Subcommittee on the Handicapped.
- 4.7 All information deriving from assessment procedures shall be considered when assessment personnel decide whether a student has a handicapping condition. The SBST or District Subcommittee on the Handicapped shall state the primary reasons for its classification and program/service category decision.
- 4.8 The SBST or the District Subcommittees on the Handicapped shall not make classification and placement decisions on the basis of the race or ethnic group of the student.
- 4.9 The SBST or the District Subcommittee on the Handicapped shall discuss the role of cultural and/or linguistic factors related to the stated referral problem or problem identified by assessment personnel, as well as other possible explanations for the student's behavior before determining if special education services are required. The substance of the discussion shall be noted in the record.
- 4.10 Before a student is classified as handicapped and in need of special education, assessment personnel shall be satisfied that there are no alternatives within the regular education system which would appropriately serve the student's needs.



Regulation of the Chancellor

Category: STUDENTS

No.: A-831

Subject: NON-DISCRIMINATORY STANDARDS AND
PROCEDURES OR THE REFERRAL,
ASSESSMENT AND PLACEMENT OF STUDENTS
WHO MAY HAVE HANDICAPPING CONDITIONS.

Page: 10 of 12

Issued: 10/1/84

- 4.11 Decisions regarding a handicapping condition shall be made only when the SBST or the District Subcommittee on the Handicapped are satisfied that the information is accurate, complete and appropriate. In reaching their decision, the assessment personnel shall make every effort to respond to parental concerns as to the accuracy, completeness and appropriateness of the information.
- 4.12 The SBST or District Subcommittee on the Handicapped should seek consensus on eligibility and placement decisions. If there is a significant dissenting opinion, the team or committee member(s) holding such opinion, may enter in the record a clear statement of the basis for the dissenting opinion(s).
- 4.13 The SBST or the District Subcommittee on the Handicapped shall not classify a student as emotionally disturbed as a result of a single incident. However, some behavior is so intense and dangerous to self and/or others that infrequent, or even single, occurrences may be of great importance.
- 4.14 In order to classify a student as handicapped, the SBST or the District Subcommittee on the Handicapped shall rely on evidence which indicates that the handicapping condition interferes with the student's ability to function effectively in a regular education placement. In all subsequent considerations of classification and/or program placement, the SBST or the District Subcommittee on the Handicapped shall consider whether the student could function effectively in a less restrictive environment.



CITY SCHOOL DISTRICT OF THE CITY OF NEW YORK

Regulation of the Chancellor

Category: STUDENTS

No.: A-831

Subject: NON-DISCRIMINATORY STANDARDS AND
PROCEDURES FOR THE REFERRAL,
ASSESSMENT AND PLACEMENT OF STUDENTS
WHO MAY HAVE HANDICAPPING CONDITIONS.

Page: 11 of 12

Issued: 10/1/84

- 4.15 The SBST or the District Subcommittee on the Handicapped shall not classify a student as emotionally disturbed based solely on the results of projective testing.
- 4.16 When a student is classified as emotionally disturbed and placed in public or private special education programs, the SBST or the District Subcommittee on the Handicapped shall have documented the evidence which indicates that a persistent pattern of maladjustive behavior interferes with the student's ability to function effectively in his or her present placement. In all subsequent considerations of classification and/or program placement, the SBST or District Subcommittee on the Handicapped shall consider whether the student could function effectively in a less restrictive environment.
- 4.17 The SBST or the District Subcommittee on the Handicapped shall make program and placement recommendations to meet the student's individual educational and related service needs in the least restrictive environment. Program and placement recommendations shall not be made on the basis of program and service availability.
- 4.18 Every SBST or District Subcommittee on the Handicapped shall develop a Phase One IEP, which provides for the educational and related service needs of the student. Every District Subcommittee on the Handicapped shall have available information about program sites which shall provide the services required by the IEP.



CITY SCHOOL DISTRICT OF THE CITY OF NEW YORK

Regulation of the Chancellor

Category: STUDENTS

No.: A-831

Subject: NON-DISCRIMINATORY STANDARDS AND
PROCEDURES FOR THE REFERRAL,
ASSESSMENT AND PLACEMENT OF STUDENTS
WHO MAY HAVE HANDICAPPING CONDITIONS.

Page: 12 of 12

Issued: 10/1/84

- 4.19 If the student is classified as emotionally disturbed, the SBST or the District Committee on the Handicapped shall specify in the student's IEP the behavioral changes which are expected as a result of special education and related services, as well as specifying academic goals to be met through special education. Recommended approaches for effecting behavioral changes shall also be included in the annual goals on the student's Phase One IEP.
- 4.20 The SBST or the District Subcommittee on the Handicapped shall document how and why the services provided by the recommended program/services category will meet the objectives of the student's IEP.

APPENDIX C

Materials for Updating the Test Resource Guide

For suggesting new and/or additional tests, submit:

Test and Manual

Comments or Special Alerts

or

Name of Author and Test

Publisher and Address

Purpose of Test

Comments or Special Alerts

For presenting new and/or additional research, submit:

Copy of research journal article, monograph, etc.

and

Citation: Journal article: author, year, title, journal
title, volume, pages

or

Book: author, year, title, place of publication,
publisher, pages

All materials should be submitted to the

Office of Educational Assessment

Test Feedback Form

For reporting positive or negative comments about particular tests included in the Test Resource Guide, submit this form to: Office of Educational Assessment.

You may want to include the following information:

- . Administration -- Too easy? Too difficult? Too lengthy?
- . Materials -- Attractive? Confusing? Out-dated? Too immature?
- . Scoring -- Easy? Difficult? Gives relevant information?
- . Use -- For formal testing? For supplemental information?

TESTS	COMMENT
-------	---------

APPENDIX D

This is to request permission to use the _____
(name of test)

in assessing _____
(name of student)

Information About Student

DOB _____ NYC ID# _____

Present placement _____

Purpose of assessment

_____ Initial _____ Reevaluation

_____ Triennial LEP: YES _____ NO _____

Information about instrument for which permission to use is sought:

1. Describe other instruments considered and found not appropriate, and why.

2. Describe proposed use of this instrument and why you believe it appropriate.

Your Signature Name of CSE Chairperson/
Clinical Supervisor Date

I approve _____/disapprove _____ of this request.

Signature of Chairperson/Clinical Supervisor

APPENDIX E

PUBLISHERS/DISTRIBUTORS DIRECTORY

Academic Therapy Publications	20 Commercial Blvd. Novato, CA 94947-6191	(415) 883-3314
Allyn & Bacon, Inc.	Rockleigh, NJ 07647	1-800-526-4799
American Association for Vocational Instructional Materials	120 Driftmier Engineering Center Athens, GA 30602	(404) 542-2586
American College Testing Program (ACT)	P.O. Box 168 Highway 1 and 180 Iowa City, IO 52240	(319) 337-1000
American Guidance Services, Inc.	Publishers Building Circle Pines, MN 55014-1796	1-800-328-2560 (203) 322-6135
A.S.I.E.P. Education Co.	3216 N.E. 27th Avenue Portland, OR. 97212	(503) 236-1317
Board of Cooperative Educational Services (BOCES) #2	Southern Westchester County 17 Berkley Drive Port Chester, NY 10573	(914) 937-3820
C.C. Publications, Inc.	P.O. Box 23699 Tigard, OR 97223	1-800-547-4800
Central Institute for the Deaf	818 South Euclid Avenue St. Louis, MO 63110	(314) 652-3200
CFKR Career Materials, Inc.	875 Haywire Ranch Road Meadow Vista, CA 95722	(916) 878-0118
Charles C. Thomas	2600 South First Springfield, IL.	(217) 789-8980
CHECpoint Systems, Inc.	1520 N. Waterman Avenue San Bernadino, CA 92404	(714) 888-3296
C.K. Conner	Director-Child Adolescent Service Human Resources Institute 227 Babcock Street Brookline, MA 02146	(617) 731-3200
Clinical Psychology Publishing Co.	4 Conant Square Brandon, VT 05733	(802) 247-6871

College Hall Press	4284 41st Street San Diego, CA 92105	
College View Printers	5040 Prescott Avenue Lincoln, NE 68529	(402) 466-6145
Consulting Psychologists Press, Inc. (CPP)	577 College Avenue Palo Alto, CA 94306	(415) 857-1444
Crane Publishing	1301 Hamilton Avenue P.O. Box 3713 Trenton, NJ 08129	(609) 586-6400
Croft Inc.	4601 York Road Baltimore, MD 21212	1-800-638-9294
CTB/McGraw Hill	Del Monte Research Park 2500 Garden Road Monterey, CA 93940	1-800-538-9547 (516) 757-6414
Curriculum Associates	5 Esquire Road No. Billerica, MA 01862-2589	1-800-225-0248 (212) 662-9364
Devereux Foundation Press	19 S. Waterloo Road Devon, PA 19333	(215) 964-3000
DLM/Teaching Resources	P.O. Box 4000 Allan, TX 75002	1-800-527-4747 (718) 317-8410
Dormac, Inc.	P.O. Box 1699 Beaverton, OR 97075-1699	1-800-547-8032 (503) 641-3128
DRP Services	The College Board 45 Columbus Avenue New York, NY 10023-6917	(212) 713-8000
Edmark Corporation	P.O. Box 3903 Bellevue, WA 98009-9990	1-800-426-0856
Educational Development Corp.	P.O. Box 45663 Tulsa, OK 74105	(918) 622-4522
Educational and Industrial Testing Service (EDITS)	P.O. Box 7234 San Diego, CA 92107	(619) 222-1666
Educational Testing Service (ETS)	Publication Order Services Roseale Road Princeton, NJ 08541-6736	(609) 921-9000

Educators Publishing Service	75 Moulton Street Cambridge, MA 02138	(617) 661-8707
Elbern Publishing	P.O. Box 09497 Columbus, Ohio 43209	(614) 235-2643
Essay Press, Inc.	P.O. Box 2323 La Jolla, CA 92038	(619) 565-6603
Ginn and Company	P.O. Box 2649 Columbus, OH 43216	(614) 876-0371
Grune and Stratton	6277 Sea Harbor Drive Orlando, FL 32887	(305) 345-4200
Harcourt, Brace, Jovanovich	757 Third Avenue New York, NY 10017	(201) 368-2200 X235
Harper and Row	Keystone Industrial Park Scanton, PA 18512	(717) 343-4761
Harvard University Press	79 Garden Street Cambridge, MA 02138	(617) 495-2600
Herbert Quay	University of Miami P.O. Box 248074 Coral Gables, FL 33124	(305) 284-2211
Human Kinetics Publishing, Inc.	Box 5076 Champaign, IL 61820	(217) 351-5076
Institute of Personality and Ability Testing (IPAT)	P.O. Box 188 Champaign, IL 61820	(217) 352-4739
Jastak Associates, Inc.	1526 Gilpin Avenue Wilmington, DE 19806	1-800-221-WRAT
Kluwer Academic Publisher	190 Old Derby Street Hingham, MA 02043	(617) 749-8321
Linguametrics Group	11 Mark Drive San Rafael, Ca 94903	(415) 499-9350
Lingui Systems, Inc.	716 17th Street Moline, IL 61265	1-800-ALL-TIME

London House Man Cons., Inc.	1550 N. Northwest Hwy. Park Ridge, IL 60068	1-800-323-5923
Mafex Assoc., Inc.	90 Cherry Street Box 519 Johnstown, PA 15902	(814) 535-3597
Marshall S. Hickey	5640 Baldwin Avenue Lincoln, NB 68507	(402) 466-6145
Martin M. Bruce Publisher	50 Larchwood Road Larchmont, NY 10538	(914) 834-1555
Modern Curriculum Press	13900 Prospect Road Cleveland, OH 44136	1-800-321-3106
National Educational Laboratory Publishers, Inc.	813 Airport Blvd. Austin, TX 78702	(512) 385-7084
NCS Interpretive Scoring Systems	P.O. Box 1416 Minneapolis, MN 55440	1-800-328-6759
Northwestern University Press	1735 Benson Avenue Evanston, IL 60201	(312) 491-5313
Office of Demographic Studies	Gallaudet College Washington, DC 20002	(202) 651-5106
Psychological Assessment Resources, Inc. (PAR)	P.O. Box 98 Odessa, FL 33556	1-800-331-TEST
Paul Kahn	1460 W 9th Stret Brooklyn, NY 11202	(718) 232-7055
Prep Inc.	1007 Whitehead Road Ext. Trenton, NJ 08638	(609) 882-2668
PRO-ED	5341 Industrial Oaks Blvd. Austin, TX 78735	(512) 892-3142
Psychological Corporation	555 Academic Court San Antonio, TX 78204	1-800-228-0752 (516) 543-6914 (202) 342-2205

Psychological Resource Services	Case Western Reserve Univ. (216) 368-3546 11220 Bellflower Road Cleveland, OH 44106	
Psychological Test Specialists	Box 1441 Missoula, MO 59801	
Psychologists and Educators, Inc.	Suite 212 211 West State Jacksonville, IL 62650	(217) 243-2135
Psychometric Affiliates	Box 3167 Munster, IN 46321	(219) 836-1661
Publishers Test Service	CTB/McGraw Hill Del Monte Research Park 2500 Garden Road Monterey, CA 93940	1-800-538-9547
Purdue University	Department 974 360 State Street W. Lafayette, IN 47906	(317) 743-9618
Richardson, Bellows, Henry and Company	1140 Conn. Avenue, N.W. Washington, DC 20036	(202) 659-3755
Riverside Publishing Co.	8420 Bryn Mawr Avenue Chicago, IL 60631	1-800-323-9540 (201) 729-6031
Sage Publishing	275 S. Beverly Drive Beverly Hills, CA 90245	(213) 274-8003
Scholastic Testing Service, Inc. (STS)	480 Meyer Road Bensenville, IL 60106	(312) 766-7150
Scott Foresman and Company	1900 East Lake Avenue Glenview, IL 60025	(312) 729-3000
Sheridan Psychological Service, Inc.	P.O. Box 6101 2710 E. Adams Avenue Orange, CA 92667	(714) 639-2595
Slosson Educational Publishing Inc.	P.O. Box 280 East Aurora, NY 14052	(716) 652-0930
SOI	343 Richmond Street El Segundo, CA 90245	(213) 322-5532
Sol Ribner	1434 East 18th Street Brooklyn, NY 11230	(718) 645-1248

SRA	155 N. Wacker Drive Chicago, IL 60606	1-800-621-0476
Stoelting Company	1350 S. Kostner Avenue Chicago, IL 60623	(312) 522-4500
Talent Assessment Inc.	4949 Sunbeam Road Jacksonville, FL 32207	(904) 731-1415
Thomas M. Achenbach	Dept. of Psychiatry(802) 656-3270 University of Vermont Burlington, VT 05401	
United Educational Services, Inc.	P.O. Box 605 E. Aurora, NY 14052	(716) 652-9131
University of Illinois Press	54 Gregory Drive Champaign, IL 61820	(217) 333-0950
University of Iowa	Publications Order Dept. Iowa City, IO	(319) 353-4171
Vocational Research Institute	2100 Arch Street Philadelphia, PA 19103	1-800-874-5387
Vocational Teacher Education	University of Northern Colorado Greeley, CO 80639	(303) 351-2939
VOKT Corporation	P.O. Box 11552 Palo Alto, CA 94306	(415) 965-4000
Walker & Company Publishers	720 5th Avenue New York, NY 10019	(212) 265-3632
Western Psychological Services	12031 Wilshire Blvd. Los Angeles, CA 90025	(213) 478-2061

BIBLIOGRAPHY

- Algozzine, B., Ysseldyke, J., & Shinn, M. (1980). Identifying children with learning disabilities: When is a discrepancy severe? University of Minnesota: IRID, Research Report No. 47.
- Ames, L.B., Learned, J., Metraux, R.W., & Walker, R.N. (1952). Child Korschach responses: Developmental trends from two to ten years. New York: Harper.
- Anastasi, A. (1976). Psychological testing. (4th ed.). NY: Macmillan.
- Anastasi, A. (1982). Psychological testing. (5th ed.). NY: Macmillan.
- Anderson, R.N., & Sisco, F.H. (1977). Standardization of the WISC-R performance, scale for deaf children. Washington, D.C.: Office of Demographic Studies, Gallaudet College, Series 7, #1.
- APA, AEHA, & NCME (1974). Standards for educational and psychological tests. Washington, DC: APA.
- APA, AEHA, & NCME (1985). Standards for educational and psychological testing. Washington, DC: APA.
- Arter, J.A., & Jenkins, J.R. (1979). Differential diagnosis-prescriptive teaching: A critical approach. Review of Educational Research, 49, 517-555.
- Beattie, R.I. (1983). Special education: A call for quality. New York City: Final Report to Mayor Edward I. Koch of the Commission on Special Education.
- Bennett, R.E. (1981). Professional competence and the assessment of exceptional children. Journal of Special Education, 15, 437-446.
- Bloom, L., & Lahey, M. (1978). Language development and language disorders. NY: Wiley.
- Bracken, B.A., McCallum, R.S., & Prasse, D.P. (in press). Peabody picture vocabulary test - revised: An appraisal and review. School Psychology Review.
- Bracken, B.A., Prasse, D.P., & Breen, M.J. (in press). Concurrent validity of the Woodcock-Johnson psychoeducational battery with regular and learning disabled students. Journal of School Psychology.
- Brown, F.G. (1980). Guidelines for test use: A commentary on the standards for educational and psychological tests. Washington, D.C.: NCME.
- Buros, O. (Ed.). (1959-1978). Mental measurements yearbooks (8th, 7th, 6th, and 5th ed.). Highland Park, New Jersey: Gryphon Press.

- Byrne, J. Hare, I. Hooper, S. Morse, B. & Sabatino. C. (1977). The role of the social history in special education evaluation. In R. Anderson, M. Freeman & R. Edwards (Eds.), School social work and PL 94-142: The education for all handicapped children act. (pp. 47-55) Silver Springs, MD: National Association of Social Workers (NASW)
- Compton, C. (1980). A guide to 65 tests for special education. Belmont, California: Pitman Learning Corp.
- Coughlin, A., & Mears, P. (1984). The social assessment of the educationally handicapped student. Silver Springs, MD: NASW, 33-46.
- Cronbach, L.J. (1970). Essentials of psychological testing (3rd ed.). NY: Harper & Row.
- Cronbach, L.J. (1983) Essentials of psychological testing. (6th ed.) NY: Harper & Row
- Doucette, J., & Freedman, R. (1980). Progress tests for the developmentally disabled: An evaluation. Cambridge, Mass.: Abt Books.
- Erickson, J.G., & Omark, D.R. (Eds.). (1981). Communication assessment of the bilingual bicultural child: Issues and guidelines. Baltimore: University Park Press.
- Exner, J. (1986) The Rorschach: A Comprehensive System: Vol 1: Basic Foundations (2nd ed.) N.Y: John Wiley
- Exner, J. (1978) The Rorschach: A Comprehensive System: Vol 2: Current research and advanced interpretations. John Wiley
- Felker, E. (1977). Systems involvement in creating a better environment for handicapped children: Parental involvement and the multidisciplinary team. In R. Anderson, M. Freeman & R. Edwards (Eds.); School social work and PL 94-143: The education for all handicapped children act. (pp. 33-46). Silver Springs, MD: NASW.
- Feuerstein, R. (1979). The dynamic assessment of retarded performers. Baltimore: University Park Press.
- Figueroa, R.A. (1979). The system of multicultural pluralistic assessment, The School Psychology Digest, 8, 28-36.
- Flegenheimer, H. (1985). Personal communication.
- Glasser, A.J., & Zimmerman, L.L. (1967). Clinical interpretation of the Wechsler intelligence scale for children. NY: Grune & Stratton.
- Goodman, J.F. (1979). Is tissue the issue? A critique of SOMPA's models and tests. The School Psychology Digest, 8, 47-62.

- Hammill, D.D., & Larsen, S.C. (1974). The effectiveness of psycholinguistic training. Exceptional Children, 41, 5-15.
- Helton, G.B., Workman, E.A., & Matuszek, P.A. (1982). Psychoeducational assessment. NY: Grune & Stratton.
- Hutt, M.L., & Briskin, G.J. (1960). The clinical use of the revised Bender-Gestalt test. NY: Grune & Stratton.
- Johnson, O.E. (1976). Tests and measurements in child development: Handbook II. San Francisco: Jossey-Bass.
- Judd, B. (1982). Consumer protection recommendations to discourage the marketing of useless and dangerous tests: A response to the test of Jastak Assessment System. Paper presented at the Annual Convention of the National Association of School Psychologists. Toronto.
- Kahn, P., & Ribner, S. (1982). Brief behavior rating scale. Psychology in the Schools, 19, 113-116.
- Kaufman, A.S. (1979). Intelligent testing with the WISC-R. NY: Wiley.
- Koppitz, E. (1963). The Bender-Gestalt test for young children. NY: Grune & Stratton.
- Koppitz, E. (1977). The visual auditory digit span test. NY: Grune & Stratton.
- Kretschmer, R.R., & Kretschmer, L.W. (1978). Language development and intervention with the hearing impaired. Baltimore: University Park Press.
- Lambert, N.M., Winomiller, M., Cole, L., & Figueroa, R.A. (1975). Standardization of a public school version of the AAMD adaptive behavior scale. Mental Retardation, 3-7.
- Lidz, C.S. (1981). Improving assessment of school children. San Francisco: Jossey-Bass.
- Malgady, R.G., Rogler, L.H., & Costantino, G. (1987). Ethnocultural and linguistic bias in mental health evaluation of Hispanics. American Psychologists, 42, 228-234.
- McDermott, P.A. (1981). Sources of error in the psychoeducational diagnosis of children. Journal of Special Education, 15, 437-446.
- McLoughlin, J., & Lewis, R. (1981). Assessing special students. Columbus, Ohio: Merrill.
- Messick, S. (1980). Test validity and the ethics of assessment. American Psychologist, 33, 1012-1027.

- Messick, S. (1984). Assessment in context: Appraising student performance in relation to instructional quality. Educational Researcher, 13, 3-8.
- Miller, J.F. (1981). Assessing language production in children: Experimental procedures. Baltimore: University Park Press.
- Mitchell, B.C. A glossary of measurement terms. Test service notebook 13. NY: Harcourt Brace Jovanovich.
- Morris, D. & Strum, I. (1986) Standard for Quality Assessment. NYCBE: unpublished manuscript.
- Mitchell, Jr., J.V. (Ed.). (1985). The ninth mental measurements yearbook (Vols. 1 and 2). Lincoln, NE: University of Nebraska
- Mulliken, R.K., & Evans, M. (Eds.) (1979). Assessment of children with low-incidence handicaps. Virginia: National Association of School Psychologists.
- Murray, E. (1980). Assessment and the individualized educational plan (IEP): Implementing assessment. In J. Murray (Ed.), Developing assessment programs for the multi-handicaped children. Springfield, IL: Charles C. Thomas.
- Murstein, B.I. (Ed.). (1965). Psychological and educational assessment of minority children. NY: Basic Books.
- NADSDE. (1978). Writing individualized assessment reports in special education: A resource manual. Washington, D.C.: NADSDE.
- Newcomer, P.L., & Hammill, D.D. (1976). Psycholinguistics in the schools. Columbus, Ohio: Charles E. Merrill.
- Nunnally, J. (1978). Psychometric theory. NY: McGraw-Hill.
- Oakland, T. (Ed.). (1977). Psychological and educational assessment of minority children. NY: Brunner/Mazel.
- Oakland, T. & Feigenbaum, D. (1980). Comparisons of the psychometric properties of the ABIC for different subgroups. Journal of School Psychology, 18, 307-316.
- Oakland, T. & Matuszek, P. (1977). Using test in nondiscriminatory assessment. In T. Oakland (Ed.) Psychological and educational assessment of minority children. (pp 52-69). NY: Brunner/Mazel
- Office of the Los Angeles County Superintendent of Schools. (1978). Pupil assessment resource guide. Los Angeles.
- Pletcher, B.S., Locks, N., Reynolds, D.S., & Sisson, B.G. (1978). A guide to assessment instructions for limited English speaking students. NY: Santillana.

Public Law 94-142, The Education for All Handicapped Children Act, 20 USC # 1401-1461.

Ramirez, M. & Castaneda, A (1984). Cultural democracy and bicultural development in education, NY: Academic Press.

Ray, S. (1979). An adaptation of the Weschsler Intelligence Scale for Children-Revised. Louisiana: Northwestern State University of Louisiana.

Regulation of the Chancellor. (1984). Non-Discriminatory Standards and Procedures for the Referral, Assessment and Placement of Students Who May Have Handicapping Conditions, A831. NY: Board of Education.

Reschly, D.J. (1980). Nonbiased assessment. Ames, Iowa: Iowa State University.

Reschly, D.J. (1984). Beyond IQ test bias: The national academy panel's analysis of minority EMR Overrepresentation. Educational Researcher, 13, 15-19.

Reeve, R.E., Hall, R.J., & Zakreski, R.S. (1979). The Woodcock-Johnson tests of cognitive ability: Concurrent validity with WISC-R. Learning Disability Journal Quarterly, 2, 63-69.

Ribner, S. (in press). A validation study of the elementary and advanced screening test of the clinical evaluation of language function.

Rosenthal, W.S. (1979). Review of the visual aural digit span test. In F.L. Darley (Ed.), Evaluation of appraisal techniques in speech and language pathology. Reading, Mass.: Addison-Wesley.

Ross-Reynolds, G. (1982). User responsibility in the use of Jastak test. Paper presented at the Annual Convention of the National Association of School Psychologists, Toronto.

Rubin, R.A. (1985). Reliability validity, and normative data for the Minnesota concise behavioral assessment tool. Independent school district #281, Robbinsdale, Minn.

Salvia, J., & Ysseldyke, J.E. (1981). Assessment in special and remedial education (2nd ed.). Boston: Houghton Mifflin.

Salvia, J., & Ysseldyke, J.E. (1985). Assessment in special and remedial education (3rd ed.). Boston: Houghton Mifflin.

Sattler, J.M. (1982). Assessment of children's intelligence and special abilities (2nd ed.). Boston: Allyn & Bacon.

Sattler, J.M., & Gwynne, J. (1982). Ethnicity and the Bender visual motor gestalt test performance. Journal of School Psychology, 20, 69-71.

- Schaefer, C., Baker, E., & Zawel, D. (1975). A factor analytic and reliability study of the Devereux elementary school behavior rating scale. Psychology in the Schools, 12, 295-300.
- Scott, R. (1981). FM: Clinically meaningful Rorschach index with minority children. Psychology in the Schools, 18, 429-433.
- SED. (1985). Guidebook for committees on the handicapped in New York State. Albany, NY: University of the State of New York, SED, Office for the Education of Children with Handicapping Conditions.
- Shepard, L.A. (1983). The role of measurement in educational policy: Lessons from the identification of learning disabilities. Educational Measurement: Issues and Practice, 14, 4-8.
- Snow, R. (1984). Placing children in special education: Some comments. Educational Researcher, 13, 12-14.
- Snyder, P.P., Snyder, R.T., & Massong, S.F. (1981). The visual motor integration test: High interjudge reliability, high potential for diagnostic error. Psychology in the Schools, 18, 55-59.
- SOMPA: A symposium. (1979). School Psychology Digest, 8, 53-61.
- Spreat, S. (1982). The AAMD behavior scale: A psychometric review. Journal of School Psychology, 20, 45-56.
- Strum, I (1985). Criteria for Determining the Quality of Assessment. NYCBE: unpublished manuscript.
- Tarczan, C. (1972). An educator's guide to psychological tests. Springfield, Illinois: Thomas.
- Teichert, M., & Daniel, P. (1980). I CRT reading: Teacher's guide. Tulsa, OK: EDC.
- Thorndike, R.L., & Hagen, E.P. (1977). Measurement and evaluation in psychology and education (4th ed). NY: John Wiley.
- Tucker, J.A. (1980). Nineteen steps for assuring nonbiased placement of students in special education. (Contract No. 400-76-00119). Washington, D.C.: National Institute of Education.
- Waddell, D.D. (1980). The Stanford-Binet: An evaluation of the technical data available since the 1972 restandardization. Journal of School Psychology, 18, 203-209.
- Weiner, I.B. Knoff, H. (Ed.) (1986). Assessing children and adolescents with the Rorschach. The assessment of child and adolescent personality (pp. 141-171). FL: Guilford press.

- Wigdor, A.K., & Garner, W.R. (Eds.). (1982). Ability testing: Uses, consequences, and controversies. Washington, D.C.: National Academy Press.
- Wilson, S. (1980). Recording guidelines for social workers. NY: Free Press.
- Winters, W. & Easton, F. (1983). The practice of school social work: An ecological perspective. New York: Free Press.
- Ysseldyke, J.E., Algozzine, B., Regan, R., & Potter, M. (1980). Technical adequacy of tests used by professionals in simulated decision making. Psychology in the Schools, 17, 202-209.
- Ysseldyke, J.E., & Regan, R.R. (1980). Nondiscriminatory assessment: A formative model. Exceptional Children, 45, 465-466.
- Ysseldyke, J.E., & Thurlow, M.L. (1984). Assessment practices in special education: Adequacy and appropriateness: Educational Psychologist, 9, 123-136.
- Zimmerman, I.L., & Woo-Sam, J.M. (1973). Clinical interpretation of the Wechsler adult intelligence scale. NY: Grune & Stratton.